DOCUMENT RESUME

ED 396 903 SE 055 572

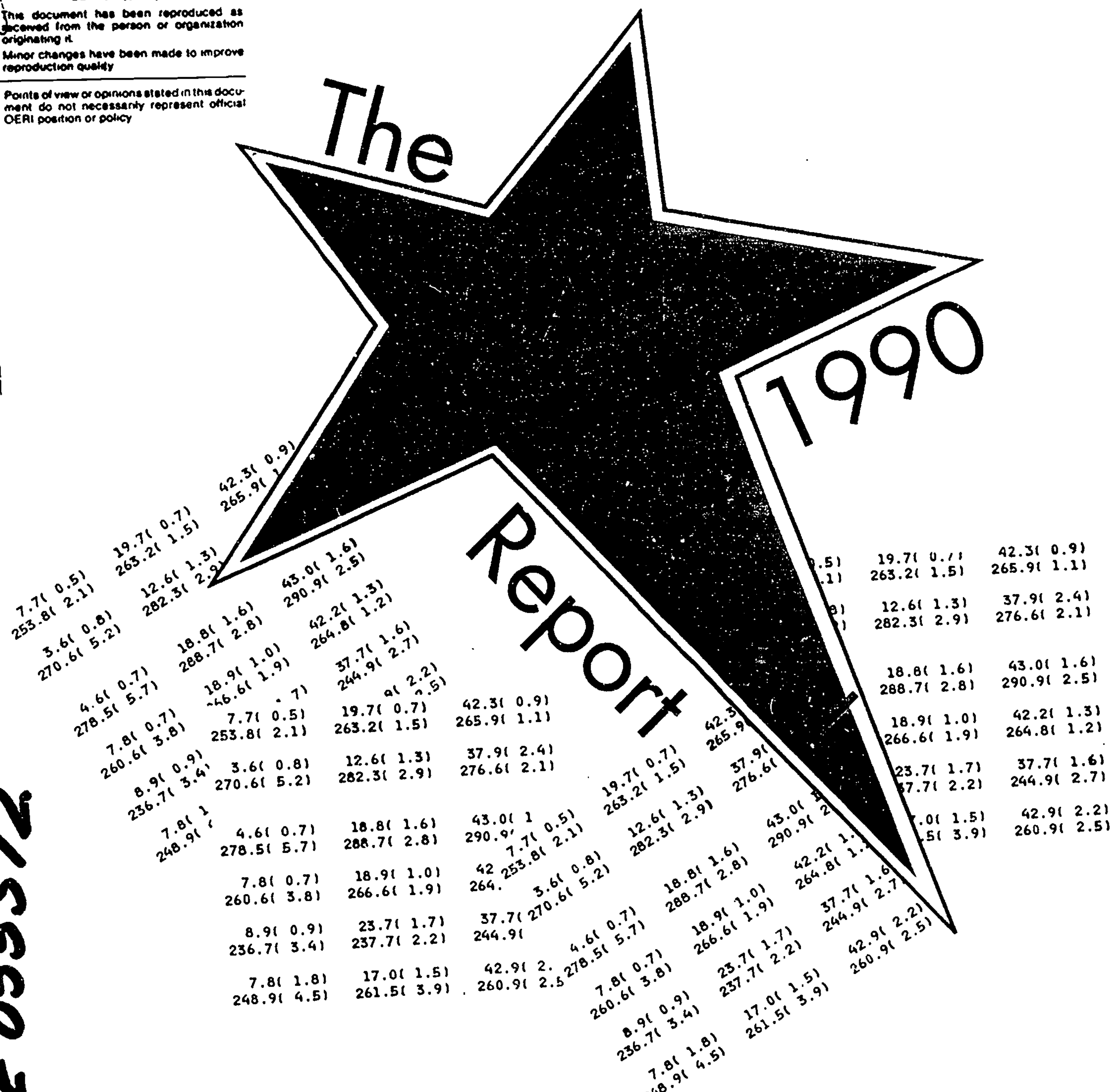
AUTHOR Johnson, Eugene G.; And Others
TITLE The NAEP 1990 Technical Report.
INSTITUTION National Assessment of Educational Progress, Princeton, NJ.
SPONS AGENCY National Center for Education Statistics (ED), Washington, DC.
REPORT NO ISBN-0-88685-123-8; NAEP-21-TR-20
PUB DATE Feb 92
NC'T 844p.
PUb ;YPE Reports - Evaluative/Feasibility (142) -- Statistical Data (110)

EDRS PRICE MF05/PC34 Plus Postage.
DESCRIPTORS *Academic Achievement; Achievement Tests; Educational Change; Elementary Secondary Education; Science Education
IDENTIFIERS *National Assessment of Educational Progress

ABSTRACT

The 1990 National Assessment of Educational Progress (NAEP) monitored the performance of students in American schools in the subject areas of reading, mathematics, science, and writing. The sample involved more than 146,000 public- and private-school students who were 9-, 13-, or 17-years old or in grades 4, 8, 11, or 12. The purpose of the report is to provide details on the instrument development, sample design, data collection, and data analysis procedures of the 1990 assessment. An additional sample of approximately 100,000 eighth-grade public-school students in 40 states and territories was assessed in mathematics as part of the 1990 Trial State Assessment. Part I, "The Design and Implementation of the 1990 NAEP" includes the following chapters: (1) "Overview of Part I: The Design and Implementation of the 1990 NAEP"; (2) "Developing the NAEP Objectives, Items, and Background Questions for the 1990 Assessments of Reading, Mathematics, and science"; (3) "Sample Design"; (4) "Assessment Instruments"; (5) "Field Operations and Data Collection"; (6) "Processing of Materials and Data"; (7) "Professional Scoring"; and (8) "Database Creation, Quality Control of Data Entry, and Database Products." Part II, "The Analysis of the 1990 NAEP Data" includes: "Overview of Part II: The Analysis of 1990 NAEP Data"; (10) "Weighting Procedures and Estimation of Sampling Variance"; (11) "Scaling Procedures"; (12) "Data Analysis for the Reading Assessment"; (13) "Data Analysis for the Mathematics Assessment"; (14) Data Analysis for the Science Assessment"; and (15) "Data Analysis for the Writing Assessment." Part III, "Statistical Summary of 1990 NAEP Data" includes: (16) "Statistical Summary of 1990 NAEP Data and Estimates of the Proficiencies of American Students." Eight appendixes provide additional information. Contains a glossary, 86 references cited in text, and an index. (ZWH)

The 1990 Report



THE NATION'S REPORT CARD naep

Eugene G. Johnson and Nancy L. Allen in collaboration with

Educational Testing Service
John R. Donoghue, John J. Ferris, Mary A. Foertsch, David S. Freund, Claudia A. Gentile, Angela Grima, Frank Jenkins, Lynn B. Jenkins, Chancey Jones, Lee R. Jones, Bruce A. Kaplan, Debra L. Kline, Stephen L. Koffler, Robert J. Mislevy, Donald A. Rock, Alfred M. Rogers, Neal Thomas, Beverly Whittington, Kentaro Yamamoto, Rebecca Zwick

Westat, Inc.
Nancy Caldwell, Dward Moore, Keith F. Rust, Renee Slobasky, James Ter Maat

National Computer Systems, Inc.
Tillie Kennel, Lavonne Mohn, Linda Reynolds, Dianne Smrdel, Brad Thayer, Lynn Zaback

What is The Nation's Report Card?

THE NATION'S REPORT CARD, the National Assessment of Educational Progress (NAEP), is the only nationally representative and continuing assessment of what America's students know and can do in various subject areas. Since 1969, assessments have been conducted periodically in reading, mathematics, science, writing, history/geography, and other fields. By making objective information on student performance available to policymakers at the national, state, and local levels, NAEP is an integral part of our nation's evaluation of the condition and progress of education. Only information related to academic achievement is collected under this program. NAEP guarantees the privacy of individual students and their families.

NAEP is a congressionally mandated project of the National Center for Education Statistics, the U.S. Department of Education. The Commissioner of Education Statistics is responsible, by law, for carrying out the NAEP project through competitive awards to qualified organizations. NAEP reports directly to the Commissioner, who is also responsible for providing continuing reviews, including validation studies and solicitation of public comment, on NAEP's conduct and usefulness.

In 1988, Congress created the National Assessment Governing Board (NAGB) to formulate policy guidelines for NAEP. The board is responsible for selecting the subject areas to be assessed, which may include adding to those specified by Congress; identifying appropriate achievement goals for each age and grade; developing assessment objectives; developing test specifications; designing the assessment methodology; developing guidelines and standards for data analysis and for reporting and disseminating results; developing standards and procedures for interstate, regional, and national comparisons; improving the form and use of the National Assessment; and ensuring that all items selected for use in the National Assessment are free from racial, cultural, gender, or regional bias.

The National Assessment Governing Board

Richard A. Boyd, Chairman
Executive Director
Martha Holden Jennings Foundation
Cleveland, Ohio

Mark D. Musick, Vice-Chairman
President
Southern Regional Education Board
Atlanta, Georgia

Phyllis Williamson Aldrich
Curriculum Coordinator
Saratoga-Warren B.O.C.E.S.
Saratoga Springs, New York

David P. Battini
High School History Teacher
Cairo-Durham High School
Cairo, New York

Parris C. Battle
Education Specialist
Dade County Public Schools
Miami, Florida

Honorable Evan Bayh
Governor
State House
Indianapolis, Indiana

Mary R. Blanton
Attorney
Blanton & Blanton
Salisbury, North Carolina

Boyd W. Boehlje
Attorney
Gaass, Klyn, & Boehlje
Pella, Iowa

Linda R. Bryant
Dean of Students
Florence Reizenstein Middle School
Pittsburgh, Pennsylvania

Honorable Michael N. Castle
Governor of Delaware
Carvel State Office Building
Wilmington, Delaware

Honorable Naomi K. Cohen
State of Connecticut
House of Representatives
Legislative Office Building
Hartford, Connecticut

Chester E. Finn, Jr.
Professor of Education and Public Policy
Vanderbilt University
Washington, D.C.

Michael Glode
Wyoming State Board of Education
Saratoga, Wyoming

William Hume
Basic American, Inc.
San Francisco, California

Christine Johnson
Director of K-12 Education
Littleton Public Schools
Littleton, Colorado

John S. Lindley
Principal
South Colby Elementary School
Port Orchard, Washington

Carl J. Moser
Director of Schools
The Lutheran Church–Missouri Synod
International Center
St. Louis, Missouri

John A. Murphy
Superintendent of Schools
Charlotte-Mecklenburg
Charlotte, North Carolina

Honorable Carolyn Pollan
Arkansas House of Representatives
Fort Smith, Arkansas

Honorable William T. Randall
Commissioner of Education
State Department of Education
Denver, Colorado

Thomas Topuzes
Senior Vice President
Valley Independent Bank
El Centro, California

Herbert J. Walberg
Professor of Education
University of Illinois
Chicago, Illinois

Diane Ravitch (Ex-Officio)
Assistant Secretary and
Counselor to the Secretary
U.S. Department of Education
Washington, D.C.

Roy Truby
Executive Director, NAGB
Washington, D.C.

NATIONAL CENTER FOR EDUCATION STATISTICS

THE NATION'S REPORT CARD naep

Eugene G. Johnson and Nancy L. Allen in collaboration with

Educational Testing Service

John R. Donoghue, John J. Ferris, Mary A. Foertsch, David S. Freund, Claudia A. Gentile, Angela Grima, Frank Jenkins, Lynn B. Jenkins, Chancey Jones, Lee R. Jones, Bruce A. Kaplan, Debra L. Kline, Stephen L. Koffler, Robert J. Mislevy, Donald A. Rock, Alfred M. Rogers, Noel Thomas, Beverly Whittington, Kentaro Yamamoto, Rebecca Zwick

Westat, Inc.

Nancy Caldwell, Dward Moore, Keith F. Rust, Renee Slobasky, James Ter Maat

National Computer Systems, Inc.

Tillie Kennel, Lavonne Mohn, Linda Reynolds, Dianne Smrdel, Brad Thayer, Lynn Zaback

February 1992
Report No. 21-TR-20

Prepared by Educational Testing Service under contract with the National Center for Education Statistics
Office of Educational Research and Improvement • U.S. Department of Education

U.S. Department of Education
Lamar Alexander
Secretary

Office of Educational Research and Improvement
Diane Ravitch
Assistant Secretary

National Center for Education Statistics
Emerson J. Elliott
Acting Commissioner

FOR MORE INFORMATION:

For ordering information on this report, write:

Education Information Branch
Office of Educational Research and Improvement
U.S. Department of Education
555 New Jersey Avenue, NW
Washington, D.C. 20208-5641

or call 1-800-424-1616 (in the Washington, D.C. metropolitan area call 202-219-1651).

Library of Congress, Catalog Card Number: 92-60172

ISBN: 0-88685-123-8

The work upon which this publication is based was performed for the National Center for Education Statistics, Office of Educational Research and Improvement, by Educational Testing Service.

Educational Testing Service is an equal opportunity, affirmative action employer.

THE NAEP 1990 TECHNICAL REPORT

CONTENTS

PART III
Statistical Summary of 1990 NAEP Data

* * *

THE NAEP 1990 TECHNICAL REPORT

LIST OF TABLES

ACKNOWLEDGMENTS

The design, development, administration, analysis, and reporting of the 1990 National Assessment of Educational Progress (NAEP) was a collaborative effort among staff from the National Assessment Governing Board (NAGB), the National Center for Education Statistics (NCES), the Council of Chief State School Officers (CCSSO), Educational Testing Service (ETS), Westat, and National Computer Systems (NCS). This report documents the NAEP design, administration, and data analysis procedures and provides insight about the rationale behind the technical decisions made. The development of this report and of the national assessment program is the result of the considerable knowledge, experience, creativity, and dedication of many individuals. It is fitting that these individuals be acknowledged for their contributions to NAEP.

The 1990 NAEP was funded through the National Center for Education Statistics in the Office of Educational Research and Improvement of the U.S. Department of Education. Emerson Elliott, acting commissioner, NCES, provided consistent support and guidance. Their staff worked closely and collegially with ETS, Westat, and NCS staff and played a crucial role in all aspects of the program. We are particularly grateful for the contributions of Gary Phillips, Eugene Owen, Stephen Gorman, Maureen Treacy, Susan Ahmed, and Raul Garza. We would also like to thank the anonymous reviewers arranged by NCES for their helpful comments. The members of the National Assessment Governing Board (NAGB) and the NAGB staff also provided advice and guidance.

ETS and NAEP management have been very supportive of NAEP's technical work. Special thanks go to Gregory Anrig and Nancy Cole as well as to Henry Braun and Charles Davis of ETS research management, and to Archie Lapointe, Ina Mullis, Jules Goodison, and David Hobson of NAEP management. Significant contributions to the project were also received from Kent Ashworth, Jay Campbell, Mary Foertsch, Claudia Gentile, Lynn Jenkins, Janet Johnson, Chancey Jones, Lee Jones, Steven Koffler, Walter MacDonald, Rebekkah Melchor-Logan, John Olson, and Beverly Whittington.

Ina Mullis deserves special recognition for her considerable influence on the past, present, and future directions of NAEP.

The guidance of the NAEP Design and Analysis Committee on the technical aspects of NAEP has been outstanding. The members are Sylvia Johnson (chair), Albert Beaton,

John Carroll, Clifford Clogg, Jeremy Finn, Bert Green, Huynh Huynh, Bengt Muthén, Ingram Olkin, Tej Pandey, and Juliet Shaffer. The untimely death of committee member T. Anne Cleary has been keenly felt. Former committee members who were also advisors for the 1990 assessment were Robert Glaser, Robert Linn, Richard Snow, and John Tukey.

The design and data analysis of the 1990 NAEP were primarily the responsibility of the NAEP research and data analysis staff with significant contributions from NAEP management, Westat, and NCS staffs. These staff members have made many innovative statistical and psychometric contributions as well as managing day-to-day data analytic operations. Major contributions were made by Nancy Allen, James Carlson, John Donoghue, Angela Grima, Frank Jenkins, John Mazzeo, Eiji Muraki, Neal Thomas, Kentaro Yamamoto, and Rebecca Zwick. Jo-lin Liang was exceptional in her role as technical assistant. Robert Mislevy and Ming-mei Wang provided valuable statistical and psychometric advice.

Special thanks go to former analysis directors Albert Beaton and Donald Rock. Dr. Beaton was the director of NAEP's statistical and psychometric activities from the beginning of ETS' participation through the design of the 1990 assessment and helped develop and direct the application of the technical procedures outlined in this and earlier technical reports. Dr. Rock was the Coordinating Director of NAEP Research through the administration and analysis of the 1990 assessment and provided leadership and support through that, the largest NAEP assessment undertaken to date.

The division of Data Analysis and Technical Research, under the outstanding leadership of John Barone, was responsible for developing the operating systems and carrying out the data analyses. Alfred Rogers and David Freund deserve special recognition for their leadership in developing and maintaining the large and complex NAEP data management systems. Alfred Rogers also deserves special mention for his role in the development of production versions of key analysis and scaling systems. Special thanks also go to David Freund, Bruce Kaplan, Edward Kulick, and John J. Ferris for their continuing roles as leaders and developers of innovative software solutions to NAEP data analysis challenges. Other members of this division who made substantial contributions of their talent, and important contributions to NAEP data analyses, were Drew Bowker, Lucie Chan, Yim Fai Fong, Steven Isham, Laura Jerry, Phillip Leung, Wing Lowe, Laura McCamley, Michael Narcowich, Jennifer Nelson, Inge Novatkoski, Craig Pizzuti, Ira Sample, and Minhwei Wang.

The staff of Westat, Inc. contributed their exceptional talents in all areas of sample design and data collection. Particular recognition is due to Renee Slobasky and Nancy Caldwell for supervising the field operations and to Keith Rust for developing and

supervising the sampling design. Morris Hansen, whose death is a tremendous loss, had considerable influence on all aspects of the assessment. Thanks are also due to Bridgett Bell, John Burke, Laurie Cameron, Fran Cohen, Mansour Fahimi, Lesly Fleming-Dunn, Dalia Kahane, Dward Moore, Marilyn Rowen, Lana Ryaboy, Debby Vivari, Mark Waksberg, Leslie Wallace, and Dianne Walsh. The exceptional data collection effort of the Westat field staff is also gratefully acknowledged.

Critical to the program was the contribution of National Computer Systems, Inc., which has been responsible for the printing, distribution, and processing of the assessment materials. The leadership roles of John O'Neill, Lynn Zaback, Dianne Smrdel, Paul McRoberts, and Linda Reynolds are especially acknowledged. Thanks go also to Tillie Kennel, Lavonne Mohn, and Brad Thayer.

Judy Alfort, Margaret Biriki, Linda Lelie, Donna Lembeck, Kay Tyberg, Marciline Yates, and Mary Varone are acknowledged for their patience and diligence in typing and proofing the many revisions of this report. Kent Ashworth was responsible for coordinating the design of the report's cover, which was created by Robin Matlack.

Nancy Allen was exceptional in guiding the direction and coordinating the writing of the analysis chapters in this report.

Special thanks go to Debra Kline, the editor of this report, for organizing, scheduling, editing, motivating, and ensuring the cohesiveness and correctness of the final report.

Finally, NAEP is grateful to the students and school staff members who participated in the assessment. Without their efforts, there would be no assessment.

Eugene G. Johnson
Coordinating Director
NAEP Research
February 14, 1992

INTRODUCTION

THE NAEP 1990 TECHNICAL REPORT

Introduction

Donald A. Rock

Educational Testing Service

The 1990 National Assessment of Educational Progress (NAEP) monitored the performance of students in American schools in the subject areas of reading, mathematics, science, and writing. The sample involved more than 146,000 public- and private-school students who were 9-, 13-, or 17 years old or in grades 4, 8, 11, or 12.

The purpose of this technical report is to provide details on the instrument development, sample design, data collection, and data analysis procedures of the 1990 assessment. Substantive results are not presented here but can be found in a series of NAEP reports on the status of and trends in student performance[1].

An additional sample of approximately 100,000 eighth-grade public-school students in 40 states and territories was assessed in mathematics as part of the 1990 Trial State Assessment. A representative sample of about 2,500 students was selected in each jurisdiction. The state-level sampling plan allowed for cross-state comparisons and comparisons with the nation in eighth-grade mathematics achievement. Technical details of the Trial State Assessment are not presented in this technical report but can be found in *The Technical Report of NAEP's 1990 Trial State Assessment Program* (Koffler, 1991).

The design and implementation of the 1990 assessment were the result of a collaborative effort of a large number of talented individuals who gave generously of their time and expertise. In addition to the staffs of Educational Testing Service, Westat, Inc., and National Computer Systems we owe a special vote of thanks to many on the staff at the National Center for Education Statistics—in particular, Emerson Elliott, Gary Phillips, Eugene Owen, and Steve Gorman—for their leadership, support, and patience in a true team effort. We were also very fortunate to have access to the accumulated wisdom of a group of acknowledged leaders in both psychometrics and statistics who served as members of the NAEP Design and Analysis Committee. Members of the present committee, which is chaired by Dr. Sylvia Johnson

[1] *The State of Mathematics Achievement: NAEP's 1990 Assessment of the Nation and the Trial Assessment of the States* (Mullis, Dossey, Owen, & Phillips, 1991); *Trends in Academic Progress: Achievement of U.S. Students in Science, 1969-70 to 1990; Mathematics, 1973 to 1990; Reading, 1971 to 1990; and Writing, 1984 to 1990* (Mullis, Dossey, Foertsch, Jones, & Gentile, 1991); *Reading in School and out of School: Students' Literary Experience and Academic Achievement from 1988 to 1990 at Grades 4, 8, and 12* (Foertsch, 1992); *The 1990 Science Report: NAEP's Assessment of Fourth, Eighth, and Twelfth Graders* (Jones, Mullis, Raizen, Weiss, & Weston, 1992); and *The Writing Students Do in School: The 1990 NAEP Portfolio Study of Fourth and Eighth Graders' School-based Writing* (Gentile, 1992).

(Howard University), include Dr. Albert E. Beaton (Boston College), Dr. John B. Carroll, Dr. T. Anne Cleary (University of Iowa), Dr. Clifford C. Clogg (Penn State University), Dr. Jeremy Finn (SUNY Buffalo), Dr. Bert F. Green, Jr. (Johns Hopkins University), Dr. Huynh Huynh (University of South Carolina), Dr. Bengt Muthén (UCLA), Dr. Ingram Olkin (Stanford University), Dr. Tej Pandey, and Dr. Juliet Shaffer (University of California at Berkeley). Former committee members who were also advisors for the 1990 assessment include Dr. Robert Glaser (University of Pittsburgh), Dr. Robert L. Linn (University of Colorado), Dr. Richard Snow (Stanford University), and Dr. John W. Tukey (Professor Emeritus, Princeton University).

An Overview of NAEP in 1990

For the 1990 assessment, as for the previous three assessments, NAEP researchers continued to build on the original design technology outlined in *A New Design for a New Era* (Messick, Beaton, & Lord, 1983). In order to maintain its links to the past and still implement innovations in measurement technology, NAEP continued its two-tiered sampling approach. Trend (or "bridge") samples use the same methodology and population definitions as in previous assessments. Cross-sectional samples use innovations associated with new NAEP technology and address current educational issues. Trend sample data are used to estimate changes in performance from previous assessments; cross-sectional sample data are used for analyses involving the current student population. In continuing to use this two-tiered approach, NAEP reaffirms its commitment to maintaining long-term trends while at the same time implementing the latest in measurement technology.

In the 1990 assessment, many of the innovations that were implemented for the first time in 1988 were continued and enhanced. For example, the use of the focused balanced incomplete block (focused-BIB) booklet design that began in 1988 was continued in 1990 for the cross-sectional samples in reading, mathematics, and science. In the focused-BIB design, an individual receives three blocks of cognitive items in the same subject area. In addition each of the blocks contains items from all of the subscales for that subject area. The focused-BIB design allows for improved estimation of composite scale scores and subscale scores within a particular subject area.

Other 1988 improvements that were used again in the 1990 cross-sectional assessment include a calendar-year definition of age for all student cohorts and the assessment of all students in two random half-samples, one in the winter and one in the spring. The availability of the winter half-sample facilitated the equating with the Trial State Assessment samples, which were also collected in the winter.

NAEP in 1990 continued to apply the plausible values approach to estimating means for demographic as well as curriculum-related subgroups. Proficiency estimates in 1990 as in 1988 were based on "draws" from a posterior distribution that was based on an optimum weighting of two sets of information—the student's responses to cognitive items and his or her demographic and associated educational process variables. This Bayesian procedure was developed by Mislevy (see Chapter 11 or Mislevy, 1991). An improvement that was implemented first in 1988 and used again in the 1990 assessment (Rogers, 1991) is the multivariate procedure that uses information from all the subscales in the estimation of the proficiency distribution on any one subscale.

An additional improvement in the plausible values technology was developed for use with the 1990 assessment. This was the use of principal components to reduce the set of specified variables in the plausible value model while capturing virtually all the variance in the student background and educational process variables that are available from the data. This approach has minimized collinearity problems and helped speed convergence.

The 1990 mathematics assessment differed from that of 1986 in age definition, time of testing, and test specifications. The 1990 NAEP cross-sectional mathematics test specifications were sufficiently different from those of 1986 to require the initiation of a new cross-sectional scale for the 1990 assessment. These specifications, which will be used again in 1992 and possibly in other future assessments, included increased emphasis on problem solving skills, the addition of estimation items, and the use of scientific calculators in the assessment.

The 1990 science cross-sectional assessment also differed from the 1986 assessment with respect to age definition and time of testing, which served to hamper comparisons of the results across the two assessments. However, because the two different assessments did share common items and reasonably similar test objectives at each age/grade level, the 1986 and 1990 scales were linked. It is anticipated that the science objectives of future NAEP cross-sectional assessments will be sufficiently different from that of 1990 to discourage linking of 1990 to succeeding assessments.

The 1990 NAEP reading assessment was linked to the 1988 assessment using a standard common IRT equating as documented in Chapter 12.

One important innovation in reporting that was initiated in 1990 was the use of Bonferroni multiple comparison procedures to form confidence intervals for the trend differences between each previous assessment year and 1990. Methods such as the Bonferroni allow one to control for the type 1 error rate for a fixed number of comparisons. In addition to the Bonferroni procedures, tests for the linear and quadratic trends were also applied to the national trend data in reading, mathematics, and science. It is anticipated that future NAEP reports will continue to build on these developments incorporating more powerful multiple comparison methods that will more optimally balance the trade-off between power and family error rates in large tables.

A major development occurring in 1990 was the implementation of the Trial State Assessment in mathematics. The Trial State Assessment, with its accompanying short reporting schedule, required a much more systematized data analysis and report writing effort than is typically found in a very complex large scale survey. The reporting on 40 states and territories led to the development of two new computer-based report writing systems that made use of artificial intelligence techniques.

In a sense, NAEP in 1990 was at a crossroads. Much of the pioneering methodology that had been developed since 1984 had steadily matured—we now needed to stand back a little bit and explore ways of making it more efficient. The 1990 assessment was carried out in the spirit of continuing the refinement of a very complex analytical system. In addition, it was a period in which there were attempts to make the system more "standardized" across subject areas. The motivation to make the system more efficient increased with the additional challenge presented by the 1990 Trial State Assessment. Because the Trial State Assessment was, as its

name indicates, a "trial," it was not clear at the outset just how feasible it would be to successfully carry out such a complex project, requiring a massive data collection, scoring, analysis, and report writing effort, in a relatively short time span. Working together, Educational Testing Service, Westat, National Computer Systems, and the National Center for Education Statistics not only accomplished all the original trial state goals within the time schedule, but along the way developed sophisticated statistical reporting systems that will set a new standard for future large-scale surveys.

Organization of the Technical Report

Part I of this report presents the details of the design of the 1990 National Assessment, summarized in Chapter 1. Other chapters describe the development of the objectives and the items used in the assessment, the sample selection procedures, the assessment booklets and questionnaires, the administration of the assessment in the field, the processing of the data from the assessment instruments into computer-readable form, and the methods used to create a complete NAEP database.

The 1990 NAEP data analysis procedures are described in Part II of the report. Chapter 9 provides a summary of the analysis steps. Subsequent chapters provide a general discussion of the weighting and variance estimation procedures used in NAEP, an overview of NAEP scaling methodology, and details of the trend and cross-sectional analyses performed for each subject area in the 1990 assessment.

Part III presents basic data from the 1990 assessment, including the properties of the measuring instruments, characteristics of the sample, and selected estimates of the proficiencies of students in each of the subject areas assessed.

PART I

The Design and Implementation of the 1990 NAEP

Chapter 1

OVERVIEW OF PART I: THE DESIGN AND IMPLEMENTATION OF THE 1990 NAEP[1]

Eugene G. Johnson

Educational Testing Service

The 1990 National Assessment collected information on the knowledge, skills. understanding, and attitudes of young Americans in reading, mathematics, science, and writing. The basis for this information was a complex sample survey involving more than 146,000 students and consisting of national samples of students who were aged 9, 13, and 17 or in grades 4, 8, 11, and 12. Additional data were collected from more than 100,000 eighth-grade public-school students from the 40 states and other jurisdictions that participated in the Trial State Assessment of mathematics.

This chapter describes the design for the 1990 assessment and gives an overview of the steps involved in its implementation, from the planning stage through the creation of edited data files. The major components of the implementation process are presented here with references to the appropriate chapters in Part I for more details. The procedures used for the analysis of the data are summarized in the overview to Part II and discussed in detail in the remaining chapters in that part of the report. Excluded from this technical report are the details of the design and analysis of the 1990 Trial State Assessment, which instead appear in *The Technical Report of NAEP's 1990 Trial State Assessment* (Koffler, 1991).

The organization of this chapter, and of Part I, is as follows:

- Section 1.1 provides an overview of the NAEP design for 1990 and describes the constituent samples. To provide background information, the section also gives the assessment schedule from the inception of NAEP in 1969 through the 1990 assessment.

- Section 1.2 summarizes the development of the objectives for each subject area in the assessment and describes the development and review of the items written to fit those objectives. Details of the objective and item development processes appear in Chapter 2.

- Section 1.3 summarizes the four-stage stratified random sampling procedures used for the 1990 assessment with a more full description provided in Chapter 3.

---

[1]The author is indebted to Albert Beaton and the authors of Chapters 2 through 8 for portions of this chapter.

- Section 1.4 discusses the assignment of the cognitive and background questions to assessment booklets and describes the focused-BIB spiral design. Chapter 4 provides a detailed description of the assessment booklets.

- Section 1.5 summarizes the field administration procedures, including the processes of training field administrators, attaining school cooperation, administering the assessment, and conducting quality control. Further details appear in Chapter 5.

- Section 1.6 describes the flow of data from the receipt of the assessment materials through data entry, validation, and resolution to the creation of edited data files. Chapter 6 provides a detailed description of the process.

- Section 1.7 discusses the professional scoring of students' responses to the open-ended items in the assessment. Details of the process are given in Chapter 7.

- Section 1.8 summarizes the creation of the database, the quality control of data entry, and lists the 1990 database products. Further details appear in Chapter 8.

1.1 THE 1990 NAEP DESIGN

Like the designs of previous assessments, the 1990 NAEP design was intended to address two occasionally competing considerations. The first was NAEP's charge to measure trends in educational achievement. The second was the need for NAEP to evolve as educational issues and priorities change over time and as new assessment technologies become available. The goal of measuring trends requires a stability in the measurement process, while the goal of keeping NAEP current requires that the measurement process be permitted to change.

The 1990 design balanced the competing demands of stability and change by basing the assessment on two distinct types of samples. The first type of sample, the trend sample, was used to estimate changes in performance from previous assessments in a subject area and used the same methodology and population definitions as in previous assessments. The second type of sample, a cross-sectional sample, was used to gather detailed information about the current student population. This sample allowed the use of new technology and population definitions and addressed new educational issues.

A number of improvements have been made in the NAEP design since ETS's first NAEP assessment in 1984. Before the 1984 assessment, NAEP used a simple matrix sampling procedure with audiotape pacing—all students in an assessment session received the same booklet of assessment items and an aurally presented stimulus was used to pace the students through the assessment items. In the 1984 assessment, balanced incomplete block (BIB) spiraling (discussed in section 1.4), which does not include aural pacing, was instituted in place of taped matrix sampling. With BIB spiraling, students in an assessment session receive different booklets resulting in a more efficient sample (as explained in section 1.4). BIB spiraling also allows the study of the interrelationships among all items included in the balanced incomplete block design. In the 1988 assessment, additional efficiencies were introduced when focused-BIB spiraling was instituted. Focused-BIB spiraling ensures that all correlations among items *within* a subject area can be estimated but, unlike the BIB designs used in 1984 and 1986,

does not require that correlations among items in *different* subject areas be estimable. This design, which was used in 1990, is simpler and more efficient than the designs used in 1984 and 1986. Like BIB spiraling, focused-BIB spiraling required the elimination of the audiotape pacing used in earlier assessments.

From its inception, NAEP has sampled students of a given age (9, 13, and 17). Since 1984, the designs have also included samples from the population of the corresponding modal grades (the school grade level of the majority of the students in a particular age group). As was the case for previous national assessments, the primary populations of inference for the 1990 assessment were in-school students of the specified ages or grades. Thus, youth of the specified age who were not attending public or private school at the time of the assessment (including, in particular, dropouts and early graduates) were excluded from the sample and from the population of inference.

To ensure that there was four years of growth between the three age/grade samples, the 1990 assessment placed all definitions of age on a calendar-year basis. Assessments prior to 1988 defined age by birth within a calendar year for ages 9 and 13 but defined the age 17 students as those who were born between October 1 of one year and September 30 of the next. Placing all age definitions on a calendar-year basis changed the modal grade for 17-year-old students from the eleventh grade to the twelfth grade. Since their age definitions were unchanged, the modal grades for ages 9 and 13 remained at grades 4 and 8.

Finally, like the 1988 assessment, all students participating in the main 1990 assessment were assessed in the winter and spring. Prior to 1988, NAEP assessed 9-year-olds in the winter, 13-year-olds in the fall, and 17-year-olds in the spring.

1.1.1 The 1990 Samples

The full 1990 assessment consisted of four types of samples: main assessment focused-BIB samples, main assessment special study samples, bridge and long-term trend samples, and the Trial State Assessment samples. A list of all assessment samples, with key characteristics, appears in Table 1-1. A description of the samples follows.

Main Assessment Focused-BIB Samples

These samples formed the basis for the cross-sectional analysis of achievement for the 1990 student population assessed in reading (labeled as [Rdg-MainP] in Table 1-1), mathematics ([Math-MainP]), and science ([Sci-MainP]). In these samples, focused-BIB spiraling (and hence printed administration) was used. The target population for the main assessment samples consisted of all students in public and private schools who belonged to one of three cohorts: students who were *either* in the fourth grade *or* 9 years old (age 9/grade 4); students who were *either* in the eighth grade *or* 13 years old (age 13/grade 8); and students who were *either* in the twelfth grade *or* 17 years old (age 17/grade 12). The main assessment represented two overlapping samples. The first sample represented students in grades 4, 8, and 12 (who could be of any age)—these are the modal grades for the students of the specified ages. The second

Table 1-1

NAEP 1990 Student Samples

Sample	Booklets	Mode	Cohort Assessed	Time of Testing	Age Definition	Modal Grade	Number Assessed
9 [Rdg-MainP]	1-7	Print	Age 9/grade 4	Winter, spring	CY	4	8,480
13 [Rdg-MainP]	1-7	Print	Age 13/grade 8	Winter, spring	CY	8	8,725
17 [Rdg-MainP]	1-7	Print	Age 17/grade 12	Winter, spring	CY	12	8,351
9 [Rdg-ABB]	8-10	Print	Age 9/grade 4	Winter, spring	CY	4	3,615
9 [Math-MainP]	11-17	Print	Age 9/grade 4	Winter, spring	CY	4	8,790
13 [Math-MainP]	8-14	Print	Age 13/grade 8	Winter, spring	CY	8	8,634
17 [Math-MainP]	8-14	Print	Age 17/grade 12	Winter, spring	CY	12	8,406
9 [Sci-MainP]	18-24	Print	Age 9/grade 4	Winter, spring	CY	4	8,418
13 [Sci-MainP]	15-21	Print	Age 13/grade 8	Winter, spring	CY	8	8,709
17 [Sci-MainP]	15-21	Print	Age 17/grade 12	Winter, spring	CY	12	8,445
9 [Math-MainT]	28	Tape	Age 9/grade 4	Winter, spring	CY	4	3,187
13 [Math-MainT]	25	Tape	Age 13/grade 8	Winter, spring	CY	8	3,182
17 [Math-MainT]	25	Tape	Age 17/grade 12	Winter, spring	CY	12	3,139
13 [Math-State]	8-14	Print	Grade 8	Winter	CY	8	100,843
9 [RW-Br84]	51-56	Print	Age 9/grade 4	Winter	CY	4	5,926
13 [RW-Br84]	51-56	Print	Age 13/grade 8	Fall	CY	8	6,233
17 [RW-Br84]	51-56	Print	Age 17/grade 11	Spring	Not CY	11	5,614
9 [RMS-Br86]	91-93	Mixed	Age 9	Winter	CY	4	6,235
13 [RMS-Br86]	91-93	Mixed	Age 13	Fall	CY	8	6,649
17 [RMS-Br86]	61-66	Print	Age 17/grade 11	Spring	Not CY	11	8,338
17 [MS-Br86]	84-85	Tape	Age 17	Spring	Not CY	11	4,411
9 [MS-BrLT]	94-95	Tape	Age 9	Winter	CY	4	4,134
13 [MS-BrLT]	94-95	Tape	Age 13	Fall	CY	8	4,455
17 [MS-BrLT]	94-95	Tape	Age 17	Spring	Not CY	11	4,402

LEGEND:

Rdg Reading
Math Mathematics
Sci Science
RW Reading and writing
RMS Reading, mathematics, and science
MS Mathematics and science

MainP Main assessment, print administration
MainT Main assessment, audiotape administration
ABB Answer book bridge (main assessment)

Br84 Bridge to 1984
Br86 Bridge to 1986
BrLT Bridge for long-term trend

Print Printed administration
Tape Audiotape administration
Mixed Mathematics and science administered by audiotape, reading administered by print

CY Calendar year: birthdates in 1980, 1976, and 1972 for ages 9, 13, and 17

Not CY Age 17 only: birthdates between Oct. 1, 1972 and Sept. 30, 1973

sample represented students of specified ages (who could be of any grade). Students were age-eligible for the main assessment if they were born in the appropriate calendar year (1980, 1976, or 1972). Only students who were attending public or private schools at the time of the assessment were included in the sample (and, therefore, in the target population). Early graduates and dropouts were excluded from the sample and the target population.

For the purposes of analysis, the age/grade samples were treated as two separate samples: (1) a representative sample of students in grades 4, 8, and 12, and (2) a representative sample of students of ages 9, 13, and 17. (A student who was both grade- and age-eligible was regarded as a member of both of these samples.) By basing analyses on a given age sample, assessment results can be related to the population of students of that age; by basing analyses on a given grade sample, assessment results can be related to the population of students in that grade.

Each age/grade sample was divided into two random half-samples, one of which was assessed in the winter and the other in the spring. The purposes of these half-samples were: (1) to allow comparison with other selected samples (assessed only in the winter or spring); and (2) to allow the study of growth in student achievement within a school year.

Main Assessment Special Study Samples

These samples were from the same populations as the main focused-BIB samples but used different assessment methodology. The special study samples in the 1990 assessment are labeled [Math-MainT] in Table 1-1. These samples were used to assess mathematics topics, such as estimation, that cannot be adequately measured using printed administration. The assessment materials for these samples were administered using audiotape pacing.

Bridge Samples

These samples are intended to adjust for changes in assessment methodology by permitting a statistical adjustment—a "bridge"—to be made. Any effect of changing the assessment methodology is bridged by comparing the responses of the bridge samples with the responses of students in other samples. There were two bridge samples in 1990.

Reading Answer Booklet Bridge: This sample of age 9/grade 4 students (labeled 9[Rdg-ABB]) was intended to bridge the results from the 1990 assessment, which used a scannable answer sheet for responses, to the results from prior assessments, where the students responded in the booklet. The age 9/grade 4 students selected for this sample were assessed for reading achievement and coded their answers in the booklet as in previous assessments. These results were compared to results from the main reading assessment. (Note that the two older cohorts received scannable answer sheets in 1988 so that no bridge was needed.)

Bridge to 1986, Age 17/Grade 11, Print Administration. This bridge (labeled 17[RMS-Br86]) consisted of a sample of age 17/grade 11 students that was comparable to the 1986 main assessment sample. These students were selected using the same age definition and were administered materials at the same time of testing as in the earlier assessment. Although the

sample definitions used for this bridge also apply to samples from the 1984 and earlier assessments, the performance of students in this bridge sample is not comparable to the students from the earlier assessments because the earlier assessments had paced audiotape administrations. The age 17/grade 11 students in this bridge were assessed in reading, mathematics, and science, using six assessment booklets first administered in 1986. These booklets contained blocks of reading, mathematics, and science items. The administration of the booklets was nonpaced and the booklets were spiraled together for administration. The effect of print versus tape administration is measured by comparing the data from this bridge with that from the *Bridge to 1986, Age 17, Tape Administration* sample (defined below).

Long-term Trend Samples

The long-term trend samples are used to estimate changes in performance from previous assessments using the same assessment methodology and population definitions as in those assessments. Because these samples were designed to link the 1990 data with data from previous assessments, they are also referred to as bridge samples. The long-term trend samples are labeled [RW-Br84], [RMS-Br86], [MS-Br86], and [MS-BrLT] in Table 1-1. Each sample was defined in the same way as an equivalent sample in a previous assessment and was administered the same assessment materials using the same procedures at the same time of year.

The target population for the long-term trend assessments consisted of all public and private school students in one of three cohorts defined using age definitions consistent with those used in 1984 and previous assessments). Students were considered age 9 and age 13 if they were born in the appropriate calendar year (1980 and 1976, respectively); students were considered age 17 if they were born between October 1, 1972 and September 30, 1973. For certain long-term trend samples, students eligible by grade were also selected. These samples consisted of students in grades 4, 8, and 11, the modal grades for the age definitions used by the bridge samples.

The long-term trend samples were directly comparable to data from previous assessments and were used to continue long-term trend lines. The various long-term trend samples, and their purposes, are:

Bridge to 1984. These long-term trend samples, labeled [RW-Br84] in Table 1-1, consisted of samples comparable to the 1984 main assessments of reading and writing. Six assessment booklets were administered at each age/grade. Each booklet consisted of at least one block of reading items and at least one block of writing items. The administration of these booklets was nonpaced (that is, no audiotape was used) and the booklets were spiraled together for administration.

Bridge to 1986, Ages 9 and 13. These are trend samples for ages 9 and 13 comparable to those used to measure trends in 1986. The samples, labeled [RMS-Br86] in Table 1-1, were collected by age only. The subject areas addressed by the [RMS-Br86] bridge were reading, mathematics, and science. Three assessment booklets, identical to those administered in 1986, were administered to each age group. Each booklet contained one block of reading, one block of mathematics, and one block of science items. As in 1986, administration of the mathematics

and science blocks was paced with an audiotape; the reading blocks were administered without an audiotape.

Bridge to 1986, Age 17, Tape Administration. This sample (labeled 17[MS-Br86]) consists of 17-year-old students. Two assessment booklets were administered to this sample. Each booklet contained blocks of mathematics and science items and was administered using a paced audiotape.

Bridge for Long-term Trend. These are labeled [MS-BrLT] and consist of age only samples for ages 9, 13, and 17. The subject areas addressed by the [MS-BrLT] bridge were mathematics and science. Two assessment booklets were administered at each age using audiotape pacing. Each booklet contained blocks of mathematics and science items where the items were last administered prior to the 1986 assessment.

Although many of the questions in the assessment booklets for the bridge and long-term trend samples also appeared in the booklets used for the main assessment, the assessment instruments were considered distinct from the main assessment instruments. Additionally, the procedures used to administer the bridge and long-term trend assessment instruments sometimes differed from those used for the main assessment. A overview of the assessment instrumentation and mode of administration appears in section 1.4, with further details presented in Chapter 4. Details of the analysis of the assessment data appear in Part II of this technical report.

The Trial State Assessment Samples

These samples, denoted by [Math-State] in Table 1-1, were samples of eighth-grade students in public schools from the states and other jurisdictions that participated in the Trial State Assessment. The assessment booklets were the same as in sample 13[Math-MainP]. Further details about the analysis of the data from the Trial State Assessment appear in Koffler (1991).

1.1.2 NAEP Assessments Since 1969

Table 1-2 shows the subject areas, grades, and ages assessed since the NAEP project began in 1969. As can be seen, besides the 1990 subject areas of reading, mathematics, science, and writing, several other subject areas have been assessed over the years—social studies, civics, U.S. history, geography, citizenship, literature, music, career development, art, and computer competence. Many subject areas have been reassessed periodically to measure trends over time.

Assessments were conducted annually through 1980, but budget restrictions since then have reduced data collection to a biennial basis. Since its inception, NAEP has assessed 9-year-olds, 13-year-olds, and in-school 17-year-olds, although the age definitions changed in 1986 and again in 1988. Because of budget restrictions, NAEP no longer routinely assesses out-of-school 17-year-olds or young adults. (A separate assessment of young adults of ages 21 to 25 was conducted in 1985 under a separate grant.)

Table 1-2
National Assessment of Educational Progress
Subject Areas, Grades, and Ages Assessed: 1969-1990

Assessment Year	Subject Area(s)	Grades/Ages Assessed										
		Grade 3	Grade 4	Age 9	Grade 7	Grade 8	Age 13	Grade 11	Grade 12	Age 17	Age 17OS*	Adult
1969-70	Science			X			X			X	X	X
1970-71	Reading			X			X			X	X	X
	Literature			X			X			X	X	X
1971-72	Music			X			X			X	X	X
	Social Studies			X			X			X	X	X
1972-73	Science			X			X			X	X	X
	Mathematics			X			X			X	X	X
1973-74	Career and Occupational Development			X			X			X	X	X
	Writing			X			X			X	X	
1974-75	Reading			X			X			X	X	
	Art			X			X			X	X	
	Index of Basic Skills									X	X	
1975-76	Citizenship/Social Studies			X			X			X	X	
	Mathematics**						X			X	X	
1976-77	Science			X			X			X		
	Basic Life Skills**									X		
	Science, Reading, Health**											X
1977-78	Mathematics			X			X			X		
	Consumer Skills**									X		
1978-79	Writing, Art, and Music			X			X			X		

* Age 17 students who had dropped out of school or had graduated prior to assessment.
** Small, special-interest assessment conducted on limited samples at specific grades or ages.

Table 1-2 (continued)
National Assessment of Educational Progress
Subject Areas, Grades, and Ages Assessed: 1969-1990

Assessment Year	Subject Area(s)	Grades/Ages Assessed										
		Grade 3	Grade 4	Age 9	Grade 7	Grade 8	Age 13	Grade 11	Grade 12	Age 17	Age 17OS*	Adult
1979-80	Reading/Literature			X			X			X	X	
	Art						X					
1981-82	Science**			X			X			X		
	Mathematics and Citizenship/Social Studies			X			X			X		
1984	Reading		X	X		X	X	X		X		
	Writing		X	X		X	X	X		X		
	Reading (long-term trend)		X	X		X	X	X		X		
	Writing (long-term trend)		X	X		X	X	X		X		
1985	Adult Literacy**											X
1986	Reading	X		X	X		X	X		X		
	Mathematics	X		X	X		X	X		X		
	Science	X		X	X		X	X		X		
	Computer Competence	X		X	X		X	X		X		
	U.S. History**							X		X		
	Literature**							X		X		
	Reading (long-term trend)		X	X		X	X	X		X		
	Mathematics (long-term trend)		X	X		X	X	X		X		
	Science (long-term trend)		X	X		X	X	X		X		

* Age 17 students who had dropped out of school or had graduated prior to assessment.
** Small, special-interest assessment conducted on limited samples at specific grades or ages.

Table 1-2 (continued)
National Assessment of Educational Progress
Subject Areas, Grades, and Ages Assessed: 1969-1990

Assessment Year	Subject Area(s)	Grades/Ages Assessed										
		Grade 3	Grade 4	Age 9	Grade 7	Grade 8	Age 13	Grade 11	Grade 12	Age 17	Age 17OS*	Adult
1988	Reading		X	X		X	X		X	X		
	Writing		X	X		X	X		X	X		
	Civics		X	X		X	X		X	X		
	U.S. History		X	X		X	X		X	X		
	Document Literacy**					X	X		X	X		
	Geography**								X	X		
	Reading (long-term trend)		X	X		X	X	X		X		
	Writing (long-term trend)		X	X		X	X	X		X		
	Mathematics (long-term trend)			X			X	X		X		
	Science (long-term trend)			X			X	X		X		
	Civics (long-term trend)						X			X		
1990	Reading		X	X		X	X		X	X		
	Mathematics		X	X		X	X		X	X		
	Science		X	X		X	X		X	X		
	Reading (long-term trend)		X	X		X	X	X		X		
	Writing (long-term trend)		X	X		X	X	X		X		
	Mathematics (long-term trend)			X			X			X		
	Science (long-term trend)			X			X			X		

* Age 17 students who had dropped out of school or had graduated prior to assessment.
** Small, special-interest assessment conducted on limited samples at specific grades or ages.

The table also indicates that in 1984, NAEP began gathering data by grade as well as by age, a practice that was continued in the 1986, 1988, and 1990 assessments. It should be noted that somewhat different age definitions were used in the 1984, 1986, and 1988 assessments. In the 1984 assessment, the younger two ages were defined on a calendar-year basis while the 17-year-olds were defined on an October 1 to September 30 basis. This resulted in modal grades of 4, 8, and 11. To allow for age cohorts that were exactly four years apart, in the 1986 main assessment, all ages were defined on an October 1 to September 30 basis, resulting in modal grades of 3, 7, and 11. Special studies (Kaplan, Beaton, Johnson, & Johnson, 1988) were conducted to measure the effect of the changes in age definition. Because of problems encountered in assessing third graders, in 1988 the ages were redefined on a calendar-year basis, with the modal grades being 4, 8, and 12. These are the age definitions used in the 1990 assessment.

1.2 DEVELOPMENT OF ASSESSMENT OBJECTIVES, ITEMS, AND BACKGROUND QUESTIONS

In 1990, NAEP conducted main assessments of students at all three age/grade levels in reading, mathematics, and science. These assessments entailed the generation of a large number of cognitive items. In addition, a large number of background and attitude questions were asked of students and information was collected from principals and teachers. Details on the item development procedures followed for the 1990 main assessment are given in Chapter 2; this section provides an overview. (In addition to the main assessment, long-term trend studies were conducted in reading, mathematics, science and writing. Since the instruments used for these studies consisted entirely of items used in previous assessments, no developmental tasks were required for their use in the 1990 assessment.)

In addition to the cognitive items, several questionnaires were developed: a common student background questionnaire to be given to all assessed students of a given age/grade, a school characteristics and policies questionnaire, teacher questionnaires for teachers of fourth- and eighth-grade students, an excluded student questionnaire, and a principal questionnaire. Each of these questionnaires was developed through a broad-based consensus process.

All items in the assessment underwent extensive reviews by subject area and measurement specialists, as well as careful scrutiny to eliminate any potential bias or lack of sensitivity to any group. Further, the items were field tested on a representative group of students. Based on the results of the field test, items were revised or modified as necessary and then again reviewed for lack of sensitivity to particular groups. With the help of staff and outside reviewers, the Item Development Panels selected the items to include in the assessment.

The assessment instruments for each subject area included both multiple-choice and open-ended items. The open-ended items were professionally scored as described in Chapter 7.

1.3 THE 1990 SAMPLE DESIGN

The sample for the 1990 NAEP assessment was selected using a complex four-stage sample design involving the sampling of students from selected schools within 97 selected

geographic areas, called primary sampling units, across the United States. The sample design was similar to that used in 1986 and 1988 and is described in detail by Westat, Inc., the firm subcontracted by ETS to select the sample, in *1990 National Assessment of Educational Progress Sampling and Weighting Procedures, Part 2: Final Report* (Rust, Burke, & Fahimi, 1992). The following sections provide an overview of each of the four stages of the sampling design with further details given in Chapter 3.

Stage 1: Primary Sampling Units

In the first stage of sampling, the United States (the 50 states and the District of Columbia) was divided into geographic primary sampling units (PSUs). Each PSU met a minimum size requirement and generally comprised either a metropolitan statistical area (MSA), a single county, or a group of contiguous counties. The PSUs were classified into four regions (Northeast, Southeast, Central, West), each containing about one-fourth of the U.S. population. In each region, PSUs were additionally classified as MSA or nonMSA. In the Southeast and West regions, the PSUs in which 20 percent of the population in the 1980 Census was either Black or Hispanic were further classified as high-minority, while the remaining PSUs in those regions were classified as not high-minority. This resulted in twelve subuniverses of PSUs.

Ninety-seven PSUs were selected for the 1990 assessment. Thirty-four PSUs were designated as certainty units because of their size, as it was cost effective to include them in the sample with certainty. Within each major stratum (subuniverse), further stratification was achieved by ordering the noncertainty PSUs according to several additional socioeconomic characteristics. Initially, 60 PSUs were selected, one per stratum from each of the noncertainty strata, with probability proportional to size (the number of school-age children from the 1980 census). To enlarge the samples of Black and Hispanic students, thereby enhancing the reliability of estimates for these groups, PSUs from the high-minority subuniverses were sampled at twice the rate of those from the other subuniverses. This was achieved by creating smaller strata within the high-minority subuniverses.

Finally, three additional PSUs were selected with the aim of decreasing the variance of estimates made from the portion of the national samples representing the aggregate of the states (plus the District of Columbia) that participated in the Trial State Assessment. All 97 PSUs were used for both the main assessments and the bridge assessments of all three age classes.[2]

[2]The term "age class" is used in this report when it is appropriate to discuss one of the three student cohorts in a general way (not necessarily in reference to a specific sample). For the 1990 assessment, age class 1 refers to age 9 and age 9/grade 4 students, age class 2 refers to age 13 and age 13/grade 8 students, and age class 3 includes the age 17, age 17/grade 11, and age 17/grade 12 students.

Stage 2: Sampling Schools

In the second stage of sampling, the public schools (including Bureau of Indian Affairs schools and Department of Defense schools) and private schools (including Catholic schools) within each of the 97 PSUs were listed according to the three age/grades. An independent sample of schools was selected separately for each of the age/grades so that some schools were selected for assessment of two age/grades, and a few were selected for all three. Schools within each PSU were selected (without replacement) with probabilities proportional to assigned measures of size with oversampling of private schools and of schools with high minority enrollment. Overall probabilities of selection for high-minority schools were twice those for other schools while the probabilities of selection of private schools were triple those of low-minority public schools of the same size. The increased probabilities of selection enlarged the samples of Black and Hispanic students and the samples of students from private schools, thereby enhancing the reliability of estimates for these groups. Details of the probabilities used for school selection appear in Chapter 3.

Schools selected for the main assessment were further classified as belonging to the winter main assessment or the spring main assessment. Random half-samples of the schools within each PSU were assigned to the winter and spring assessments.

The samples of schools for the bridge assessments were drawn in a manner very similar to that used for the main assessments. The chief difference in the two samples was that private schools and schools with high minority enrollment were not oversampled for the bridge assessments.

The overall school cooperation rate exceeded 80 percent at each age/grade. In certain instances, refusing schools were replaced by substitutes according to the rules indicated in Chapter 3.

Stage 3: Assigning Assessment Sessions to Schools

In the third stage of sampling, assessment sessions were assigned to the sampled schools, as described in Chapter 3. An assessment session typically consisted of 25 to 30 students, all of whom could be assessed following the same procedures. There were two general types of sessions in the 1990 assessment: (1) tape sessions, where every student was administered the same booklet and where audiotape prompts paced the students through at least part of the booklet, and (2) print sessions, where a number of distinct booklets were administered and where no audiotape pacing was used. (Print sessions are also called spiral sessions, since the assessment booklets were spiraled for administration—see section 1.4.1.) The assignment of sessions to schools was designed to maximize the number of session types conducted within each PSU, where each session type corresponded to a separate sample of the population of students.

Stage 4: Sampling Students

In the fourth stage of sampling, a consolidated list was prepared for each school of all grade-eligible and age-eligible students for the age class for which the school was selected. To

provide the target sample size, a systematic selection of eligible students was made from this list, if necessary. In small and medium-sized schools all eligible students were in the sample. For schools assigned to more than a single session type, students were assigned by Westat district supervisors to one of the various session types using specified procedures. A student was not assigned to more than one session.

Stage 4a: Excluded Students

It is NAEP's intention to assess all selected students. However, certain selected students may be judged by school authorities as being incapable of participating meaningfully in the assessment. For each of these students, school staff completed an excluded student questionnaire, listing the reason for exclusion and providing some background information.

Specific guidelines for exclusion were provided for all samples in the 1990 assessment. However, somewhat different criteria were used for the long-term trend samples than for the main assessment samples. The exclusion guidelines for the long-term trend samples were the same as those used in previous assessments. Three types of students could be excluded under these guidelines—non-English speaking students, educable mentally retarded students who were judged incapable of meaningfully responding to exercises appropriate to their age level, and students so functionally disabled that they could not perform in the NAEP assessment situation.

The criteria used for the main assessments provided more specific rules for exclusion. A student identified as having Limited English Proficiency (LEP) could be excluded if the student was a native speaker of a language other than English, had been enrolled in an English-speaking school for less than two years, and was judged incapable of taking part in the assessment. Students to be excluded for reasons other than LEP had to be special education students with Individualized Education Plans (IEP), or equivalent designation, who were mainstreamed less than 50 percent of the time in academic subjects and/or were considered unassessable by the IEP team.

Stage 4b: Sampling Teachers

The mathematics teachers of fourth-grade and eighth-grade students sampled for the main assessment of mathematics and the science teachers of eighth-grade students sampled for the main assessment of science were identified and asked to complete a questionnaire (described in Chapter 4) about their background and experiences and about instructional practices, by class, for any classes containing assessed students.

Stage 4c: The School and Principal Questionnaires

A school characteristics and policies questionnaire was mailed to every sampled school by Westat before the assessment. The Westat supervisor then collected the questionnaires and returned them to ETS. The school characteristics and policies questionnaire is described in Chapter 4. The principal questionnaire, distributed to the principal of each sampled school by

Westat before the assessment, was used to estimate the number of age/grade-eligible students and to determine the correct "size and type of community" classification for each school.

1.4 ASSESSMENT INSTRUMENTS

Five types of instruments were used in the 1990 assessment: student assessment booklets (which included the student common background questionnaire), excluded student questionnaires, teacher questionnaires, a principal questionnaire, and a school characteristics and policies questionnaire. This section provides an overview of these instruments; more detailed information can be found in Chapter 4.

1.4.1 Student Assessment Booklets—Main Assessment

The student assessment booklets for the main assessment contained both cognitive and noncognitive items. The total testing time was approximately 45 minutes for age 9/grade 4 students and 56 minutes for students at the older grades and ages. A block of common background questions appeared first in every booklet and required 10 minutes for age 9/grade 4 students and 6 minutes for the older students. This block was followed by a 5-minute block of subject-specific background questions and (typically) three 15-minute blocks of cognitive items (10-minute blocks for age 9/grade 4).

The assembly of cognitive items into booklets and their subsequent assignment to assessed students was determined by a *balanced incomplete block* (BIB) design with *spiraled* administration. The first step in implementing BIB spiraling was to divide the items within a subject area into units called blocks, where each block was designed to take 15 minutes for the older students to complete. For the age 9/grade 4 students, blocks were created that required 10 minutes for completion. Some blocks were administered at more than one age/grade; additional items were added to the end of age 9/grade 4 blocks that were also administered to older students.

Using a partially balanced incomplete block design, the blocks were then assembled into booklets, such that each booklet contained a set of background questions and three blocks of cognitive questions. In a completely balanced incomplete block design, the subject area blocks would be assigned to booklets in such a way that each block appears in the same number of booklets and every pair of blocks appears together in exactly one booklet. This is the *balanced* part of the method; the *incomplete* part refers to the fact that no booklet contains all items and hence incomplete data is yielded for each assessed student. Such a design allows the computation of the correlation between each pair of items but generates a vast number of different booklets, particularly if blocks from different subject areas are to be paired.

In 1990, the BIB design was focused—that is, each block of items within each subject area was paired with every other block within that subject area but not with blocks of items from other subject areas. The focused-BIB design used in 1990 called for seven blocks of cognitive items in a given subject area at a given age/grade to be assembled into seven booklets, where each pair of the seven blocks appears in exactly one booklet and where each block appears in three booklets. The focused-BIB design also balances the order of presentation of the blocks of

items so that every block of items appears as the first cognitive block in one booklet, as the second block in another booklet, and as the third and last block in a third booklet.

An example of the focused-BIB design with seven blocks (labeled A, B, C, D, E, F, G) and seven booklets (labeled B1 through B7) is as follows:

Booklet	Blocks		
B1	A	B	D
B2	B	C	E
B3	C	D	F
B4	D	E	G
B5	E	F	A
B6	F	G	B
B7	G	A	C

In addition to the focused-BIB booklets, a special booklet was created as a part of the mathematics assessment for each age/grade. This booklet consisted of three blocks measuring estimation and complex problem-solving abilities, neither of which can be adequately measured with printed administration. Accordingly, this booklet was administered using audiotape pacing.

A total of 25 different booklets were assembled for age 9/grade 4, 22 different booklets for age 13/grade 8, and 22 different booklets for age 17/grade 12. These booklets were then spiraled and placed into bundles. Spiraling involves interleaving the booklets in regular (systematic) sequence so that each booklet appears an equivalent number of times in the sample. Booklets were packaged together in bundles of 25 to 27 booklets, which were large enough to accommodate a typical assessment session. The bundles were designed so that each booklet would appear equally often in each position in a bundle. The reading and science booklets were bundled together. Because mathematics was assessed in separate sessions from reading and science, due to the use of calculators, the mathematics booklets were bundled separately.

The final step in the BIB-spiraling procedure was the assignment of booklets to the assessed students. The students within an assessment session were assigned booklets in the order in which the booklets were bundled. The result was that, typically, each student in an assessment session received a different booklet and, even in schools with multiple sessions, only a few students received the same booklet or block of items. In the 1990 BIB-spiral design, representative and randomly equivalent samples of about 2,600 grade- or age-eligible students responded to each item (resulting in samples of about 2,000 students eligible by age and 2,000 eligible by grade).

BIB spiraling differs from the simpler matrix sampling scheme used by NAEP prior to 1984 to assign items to students. In the earlier scheme, the pool of items was divided into distinct booklets requiring about 45 minutes to administer, and all students within an assessment session were given the same booklet. Because all students in a session received the same booklet, it was possible to accompany the administration with a paced audiotape of the exercise stimuli, with the aim of minimizing the effect of a student's reading ability on performance in other subject areas. However, since each item appeared in a single booklet, it was impossible to estimate correlations between items appearing in different booklets. Furthermore, the

administration of the same items to clusters of students within schools results in an increase in sampling variability over an unclustered sample of the same size because of intracluster correlation.

The BIB-spiral design permits the estimation of correlations between all items within a subject area. Furthermore, since the spiral design presents each block of items to fewer persons in any single school, but to more schools, than the simple matrix sampling design, the cluster effect is markedly reduced, leading to a sample with high statistical efficiency. The spiral design does preclude the use of audiotape pacing. Since each student within a session responds to a different set of items, the instructions and the items themselves must be read by the student as audiotape administration would be unmanageable. For this reason, simple matrix sampling designs, where every student receives the same booklet, are used in the special cases requiring audiotape pacing, such as the assessment of mathematics estimation and higher-order thinking skills.

1.4.2 Student Assessment Booklets—Bridge and Long-term Trend Samples

There were five distinct bridge and long-term trend samples in the 1990 assessment, each of which required the use of special booklets:

Bridge to 1984. Six booklets were used at each of the three age/grade classes. These booklets were identical to booklets used in the 1984 assessments of reading and writing. Each booklet consisted of a common background block and three cognitive blocks, either two reading and one writing or one reading and two writing. All cognitive blocks also contained subject-related background questions. The booklets were administered without audiotape and were spiraled through the assessment session.

Bridge to 1986 for Ages 9 and 13. Three booklets were used for each of the ages 9 and 13. These booklets were identical to booklets administered in 1986 and contained one block each of reading, mathematics, and science items in addition to a common background block. All cognitive blocks also contained subject-related background questions. In each session, all students were administered the same booklet. The mathematics and science blocks were paced with an audiotape; the reading blocks were administered without an audiotape.

Bridge to 1986 for Age 17/Grade 11, Print Administration. Six assessment booklets, consisting of blocks of reading, mathematics, and science items, were used for administration to age 17/grade 11 students. These booklets were identical to booklets administered in 1986 and were given without audiotape pacing. All booklets were administered to students in the same assessment session.

Bridge to 1986 for Age 17, Tape Administration. Two booklets, identical to booklets administered in 1986, were administered to age 17 students. Each booklet contained background questions and three blocks of cognitive items, either two mathematics and one science or one mathematics and two science. All students within a session were administered the same booklet. The administration of each booklet was paced by an audiotape.

Bridge for Long-term Trend. Two booklets were created for each of the ages 9, 13, and 17. Each booklet consisted of background questions and either two blocks of mathematics items and one block of science items or one block of mathematics and two blocks of science items. All cognitive items in these booklets had last been administered in an assessment prior to 1986. In each session, all students were administered the same booklet. The administration of each booklet was paced by an audiotape.

1.4.3 Other Instruments

In addition to the student assessment booklets, four other instruments were used to gather data as a part of the assessment:

Teacher questionnaires were administered to the mathematics teachers of fourth- and eighth-grade students who participated in the main mathematics assessment and to the science teachers of eighth-grade students who participated in the main science assessment. These questionnaires were designed to gather information about the characteristics of the teachers of assessed students and about the curricula and teaching methods used in classrooms across the country.

School characteristics and policies questionnaires were completed by school principals or their representatives, who provided information about school administration, staffing patterns, special programs, subject requirements, and school resources.

Excluded student questionnaires were completed by school personnel for each sampled student excluded from the assessment. The questionnaires requested information about the reasons for exclusion as well as basic demographic characteristics of the student.

Principal questionnaires, distributed to the principal of each sampled school before the assessment, was used to estimate the number of age/grade-eligible students and to determine the correct "size and type of community" classification for each school.

1.5 FIELD OPERATIONS AND DATA COLLECTION

Field operations and data collection for the 1990 assessment were the responsibility of Westat, Inc., and are documented in Chapter 5. The field operation was conducted by a staff at Westat's home office and a larger staff in the field. The Westat home office staff coordinated all activities related to field operations and managed materials distribution and home-office receipt of assessment reporting forms. The field staff consisted of area supervisor, assessment supervisors, and exercise administrators. The assessment supervisors, who were trained by Westat, were each responsible for the assessment activities in one or more PSUs. Although ETS made initial contact with participating school districts, each assessment supervisor was primarily responsible for making follow-up contacts with these districts, recruiting and training exercise administrators to work with them in administering the assessment sessions, arranging the assessment sessions, and selecting the sample of students to be assessed within each school. The assessment supervisors and the exercise administrators administered the assessments, filled

out the necessary forms, performed process control, and shipped the assessment booklets and forms to National Computer Systems (NCS), the subcontractor responsible for processing NAEP materials and data.

Gaining school cooperation was the joint responsibility of Westat and ETS staff. ETS made the preliminary contacts preparatory to obtaining school cooperation by first contacting the Chief State School Officers, informing them that schools within their states had been selected for the assessment and, in a later letter, listing the selected schools and districts. Later mailings were sent to superintendents of public schools and parochial schools and principals of private schools for all schools selected in the assessment. These materials provided an explanation of NAEP, a list of the selected schools in the official's jurisdiction, and a cover letter explaining that a Westat district supervisor would contact them to set up an introductory meeting. Westat district supervisors then scheduled and conducted introductory meetings, worked with the schools to schedule the assessments, and, with the exercise administrators, conducted the assessments. The overall cooperation rate of schools originally selected for all phases of the 1990 assessment was 86 percent. Further detail on school participation rates is given in sections 3.2 and 3.3 of Chapter 3.

The main assessment sessions were conducted between January 8 and May 18, 1990 at all age/grade levels. The winter portion of the main assessment and the age 9/grade 4 bridge assessments were carried out between January 8 and March 16; the spring portion of the main assessment and the age 17/grade 11 bridge samples were conducted between March 19 and May 18, 1990. The age 13/grade 8 bridge assessments were carried out between October 9 and December 13, 1989.

An automated management system tracked and recorded the progress of field work throughout the 1990 assessment period. In addition, progress was constantly monitored through telephone reports held between the area supervisors and the assessment supervisors and between the area supervisors and the home office staff.

Both Westat and ETS participated in the quality control of the field administration. The quality control involved on-site visits by Westat and ETS staff to verify the sampling of the students and to observe the conduct of the assessment by the supervisors and the exercise administrators.

1.6 MATERIALS AND DATA PROCESSING

After completing an assessment session, Westat field supervisors and exercise administrators shipped the assessment booklets and forms from the field to National Computer Systems for entry into computer files, professional scoring, checking, and creating the data files for transmittal to ETS. Careful checking assured that all data from the field were received. More than 280,000 booklets and questionnaires were received and processed for the 1990 assessment. The extensive processing of these data is detailed in Chapter 6.

The student data were transcribed into machine-readable form by scanning the student instruments with an optical scanning machine. An intelligent data entry system was used for resolution of the scanned data, the entry of documents rejected by the scanning machine, and

the entry of information from the questionnaires. Additionally, each piece of input data was checked to verify that it was of an acceptable type, that it was within a specified range or ranges of values, and that it was consistent with other data values. The entry and editing of materials is discussed in Chapter 6.

1.7 PROFESSIONAL SCORING

Items requiring a written response from the student (open-ended items) were included in the main assessments in reading, mathematics, and science, the Trial State Assessment in mathematics, and the bridge assessments in reading, mathematics, and writing. More than three million open-ended item responses were read and marked by the professional scoring staff for the 1990 assessments.

Chapter 7 describes the professional scoring operation, including an overview of the scoring guides, the training procedures, and the scoring process for each subject area.

1.8 CREATION OF THE DATABASE

Before any analyses could begin, the student response data, school, teacher, and excluded student questionnaire data, and all sampling weights had to be integrated into a coherent and comprehensive database. This database was used for all analyses. The database was also the source for the creation of three NAEP database products: the integrated information database, the restricted-used data files, and the secondary-use data files. The quality of the data resulting from the complete data entry system, from the actual instruments collected in the field to the final machine-readable database used in analysis was verified by selecting field instruments at random and performing a character-by-character comparison of these instruments with their representations in the final database. Chapter 8 provides details on the database, quality control activities, and database products.

Chapter 2

DEVELOPING THE NAEP OBJECTIVES, ITEMS, AND BACKGROUND QUESTIONS FOR THE 1990 ASSESSMENTS OF READING, MATHEMATICS, AND SCIENCE

Mary A. Foertsch, Lee R. Jones, Stephen L. Koffler

Educational Testing Service

The subject areas constituting the 1990 assessments were specified by the 1988 legislation authorizing the National Assessment of Educational Progress and included reading, mathematics, and science. In addition, for the first time in its history, the NAEP in 1990 included a Trial State Assessment Program in which a representative sample of eighth-grade public-school students in 37 participating states, the District of Columbia, and two territories were assessed in mathematics. A summary of each assessment area follows:

Reading: A reading assessment, developed in 1988, was updated for 1990 and included questions designed to measure reading as a process involving the construction and examination of meaning. Students at grades 4, 8, and 12 were asked multiple-choice as well as a few open-ended questions about literary and informational passages and about documents.

Mathematics: For the nation, a newly developed mathematics assessment was administered at grades 4, 8, and 12 that included the use of four-function calculators at grade 4, scientific calculators at grades 8 and 12, and open-ended problem solving questions at all grades. In addition, estimation and complex problem solving were assessed in a special study using audiotapes that paced students through the questions. For the 40 participating states and jurisdictions in the Trial State Assessment, the newly developed mathematics assessment was administered at grade 8.

Science: A newly developed science assessment at grades 4, 8, and 12 included two types of open-ended questions that asked students to write brief responses demonstrating their ability to conduct scientific inquiry and to draw illustrations indicating their grasp of scientific events.

From its inception, NAEP has developed assessments through a consensus process and the 1990 assessment was no exception. Educators, scholars, and citizens representative of many diverse constituencies and points of view designed objectives for each of the three subject areas, proposing goals they felt students should achieve in the course of their education. After careful review, the objectives were given to item writers who were experts in their respective field and who developed assessment questions appropriate to the objectives. All questions underwent extensive reviews by subject-matter, measurement, and bias/sensitivity specialists. They were assembled and printed into booklets suitable for matrix sampling and then administered either

by a trained field staff (for the national program) or by state or local school district staff (for the Trial State Assessment Program) to stratified, multistage probability samples of students.

All 1990 development efforts were governed by four major considerations:

1) As specified in the legislation, the objectives were to be developed through a consensus process involving subject matter experts, school administrators, teachers, and parents, and the items were to be carefully reviewed for potential bias.

2) As outlined in the ETS proposal for the administration of the NAEP contract, the development of objectives and items for each subject area was to be guided by a Assessment Development Panel.

3) As described in the *ETS Standards for Quality and Fairness* (Educational Testing Service, 1987), all materials developed at ETS were to be in compliance with specified procedures.

4) As required by federal regulations, all NAEP items were to be submitted to a complex clearance process.

The development effort for the 1990 assessment included questionnaires for students, teachers, and school administrators, in addition to a substantial number of cognitive items for each of the three subject areas.

The following sections include general overviews about setting objectives and developing items and specific details about developing the objectives and the assessments. A list of the consultants who participated in the 1990 development process is included in Appendix A.

2.1 GENERAL OVERVIEW OF THE 1990 OBJECTIVES

The subject-area objectives for each NAEP assessment are determined through a legislatively mandated consensus process managed by NAEP's governing board (previously, the Assessment Policy Committee, or APC; now the National Assessment Governing Board, or NAGB). These objectives typically take the form of frameworks or matrices delineating the important content and process areas to be assessed. The various frameworks for the 1990 assessments are described below and discussed in detail by Mullis (1990).

The 1990 **reading** assessment was used to measure short-term trend from 1988 and was structured to examine comprehension as it occurs in two modes of reading: "Reading to Construct Meaning" and "Reading to Examine or Extend Meaning." Although both modes of reading are intertwined in most actual reading experiences, they were separated in the assessment to clarify the distinction between them. As with previous NAEP reading assessments, the 1990 reading assessment measured students' ability to read based on a variety of passages, ranging from textbook materials, documents, and news articles to poems, essays and stories. Passages were grouped into three categories—literary text, informational text, and documents—because these categories represent the types of materials that students commonly encounter in and out of school and are expected to be capable of reading.

In reading, as in other subject areas, NAEP addressed the perennial questions of what to assess and how to so through a consensus process involving curriculum specialists, teachers, school administrators, researchers, parents, concerned citizens, public officials, and business leaders. For the 1990 reading assessment, NAEP's Assessment Development Panel reviewed the structure of the 1988 reading assessment and recommended certain modifications. In particular, efforts were made to integrate new theory and research on the learning and teaching of reading and to reflect innovative approaches to reading assessments developed in Michigan and Illinois as part of their state assessment programs (Michigan State Board of Education, 1987; Illinois State Board of Education, 1988).

Throughout the development process, the reading objectives were reviewed by members of the Assessment Development Panel and Item Development Panel and by outside consultants representing various constituencies, including members of professional organizations, teachers, school officials, and interested lay persons. While objectives resulting from such a consensus process reflect neither a narrowly defined theoretical framework nor every view of every participant, they do represent the thinking of a broad cross-section of individuals who are expert in the areas of literacy research and reading instruction and who are deeply committed to the improvement of reading in our schools.

NAEP's 1990 **mathematics** assessment framework is a five by three matrix specifying five content areas (numbers and operations; measurement; geometry; data analysis, statistics, and probability; and algebra and functions) and three process or ability areas (conceptual understanding, procedural knowledge, and problem solving).

The mathematics objectives were developed under the auspices of the Council of Chief State School Officers (CCSSO) through a special NAEP Planning Project sponsored by the National Center for Education Statistics (NCES) and the National Science Foundation. This project involved widespread participation and review by many groups, including an objectives committee of mathematics educators; a steering committee of 18 members representing policy makers, practitioners, and citizens at large; distribution to the mathematics supervisor in each state education agency for review by state committees; and reviews by mathematics scholars, NCES, and NAEP's governing board.

The 1990 **science** framework encompasses a matrix of three content areas—life sciences, physical sciences, and earth and space sciences—and three broad process areas or thinking skills—conducting inquiries, solving problems, and knowing science. The foundation of the matrix is understanding the nature of science.

Objectives for the 1990 science assessment were developed under a grant from NCES to ETS, which managed the consensus process. An Item Development Panel including science education and assessment experts from universities and professional associations, as well as from state and local education agencies, guided the process with the assistance of ETS staff. An iterative series of reviews was conducted—by the Item Development Panel, science education and assessment experts, science curriculum coordinators from schools and state education agencies, scientists, school administrators, and the APC.

2.2 GENERAL OVERVIEW OF PROCEDURES FOR DEVELOPING THE ITEMS

A carefully developed and tested series of steps, similar to those for past NAEP assessments, were used to create assessment items that reflected the mathematics, reading, and science objectives and that measured achievement related to them (see sections 2.3—2.5 for more detail). The steps were as follows:

1) Each Item Development Panel and the NAEP staff determined what specific aspects of the objectives could be measured given the realistic constraints of resources and the feasibility of implementing the measurement technology. Each respective committee made recommendations about priorities for the assessment and types of items to be developed. The measures were constrained to those that could be administered via paper and pencil technology to groups of students.

2) The existing pool of items to be used to measure change from previous assessments (trend items) was reviewed in detail and trend items selected in all subject areas.

3) Item specifications were then developed and prototype items were created to reflect the type of questions that had been suggested.

4) Item writers with subject-matter expertise and skills and experience in creating items according to specifications were identified from both inside and outside ETS and scheduled for item development tasks.

5) Newly created items were reviewed and revised by staff and external reviewers, including the Item Development Committee. The items for the eighth-grade mathematics assessment were also reviewed by representatives from the State Education Agencies.

6) Further language editing and sensitivity reviews were conducted as required by the ETS standards.

7) Field test materials were prepared, including the materials necessary to secure Office of Management and Budget (OMB) clearance.

8) The field test for the national program was conducted with a representative group of students from across the country. The Trial State Assessment Program, the field test of the eighth-grade mathematics assessment was conducted in 24 states, the District of Columbia, and three territories.

9) Field test booklets were scored and the results analyzed.

10) Based on these analyses and the results of the field test, items were revised or modified and re-edited. They once again went through the required ETS sensitivity review.

11) With the help of staff and outside reviewers (including state representatives for the eighth-grade mathematics assessment), each Assessment Development Panel selected the items to include in the assessments.

12) Items were assembled into "blocks" (15-minute mini-tests) with attention given to balancing content coverage and difficulty levels.

13) After a final review and check to ensure that each assessment booklet and each block therein met the overall guidelines for the assessment, the booklets were typeset and printed.

The following sections describe the development of the reading, mathematics, and science assessments in more detail.

2.3 DEVELOPING THE READING ASSESSMENT

2.3.1 Reading Objectives

The objectives for the 1990 reading assessment reflect an effort to integrate new theory and research on the learning and teaching of reading, and to reflect innovative approaches to the assessment of reading. Central to the 1990 reading objectives is the conviction that pedagogy and assessment should be based on an interactive view of reading in which the reading process—and, therefore, comprehension—is influenced by a range of factors related to the nature of the text, the purposes for reading, and the reader.

The 1990 reading assessment was structured to examine comprehension as it occurs in two modes of reading: "Reading to Construct Meaning" and "Reading to Examine or Extend Meaning." Although both modes of reading are intertwined in most actual reading experiences, they were separated in the assessment to clarify the distinction between them.

In the "Reading to Construct Meaning" mode, readers direct their efforts towards building a general model of the text's meaning and significance based on their expectations, existing knowledge and perceptions of the new information encountered during the reading process. Their primary purposes are to find the gist of the author's message, capture details of personal interest or immediate importance, review major themes and main ideas, recognize similarities or differences with their own ideas or other texts they have read, or evaluate the text's potential to provide opportunities for learning or enjoyment. Alternatively, individuals may read to ascertain general linkages among events in a story, a historical account, or a biological process.

In "Reading to Examine Meaning," readers try to broaden and deepen their comprehension of the text by filling in details that embellish their general understanding, explore relations among ideas that are not immediately apparent, and use their existing knowledge to establish new connections with ideas from the text. Readers may read for nuances to predict outcomes, infer links in a causal chain of reasoning, evaluate the text according to explicit or implicit criteria, or develop and test their inferences.

In document reading, the two modes of comprehension are "Locate or Compare Information" and "Evaluate Information," respectively, to reflect the different strategies involved in reading documents. Essentially, the two document reading modes differ from those in informational and literary text because they place less emphasis on print and more emphasis on graphic elements.

2.3.2 Reading Materials

As with previous NAEP reading assessments, the 1990 reading assessment measured students' ability to read based on a variety of passages, ranging from textbook materials, documents, and news articles to poems, essays and stories. Passages were grouped into three categories—literary text, informational text, and documents—because these categories represent the types of materials that students commonly encounter in and out of school and are expected to be capable of reading. The reading passages within each text category were derived from a range of sources, including trade books, school texts, children's literature, newspapers and magazines, and reference books. Final selection of a passage was guided by the following criteria: concurrence with the assessment objectives; richness, accuracy, and difficulty level of the content; quality of the writing; and presentation.

2.3.3 Distribution of Assessment Items

The distribution of assessment items for the 1990 reading assessment was designed to reflect the association between two dimensions of reading highlighted in the framework: Type of Text and Mode of Comprehension.

The percentages of items by mode of comprehension was the same at each grade level: Seventy percent of the assessment was intended to be devoted to constructing meaning, and 30 percent to extending meaning. However, the percentage of items by type of text varied by grade level, as shown in Table 2-1.

Table 2-1
Percentage Distribution of Reading Items by Type of Text and Grade

	Grade		
Type of Text	4	8	12
Information Text	42%	73%	67%
Literary Text	52%	17%	21%
Documents	6%	10%	12%

Tables 2-2, 2-3, and 2-4 provide a description of the number of items for each text type and mode of comprehension that was included at each grade level for the main BIB design.

Table 2-2
Text Type by Mode of Comprehension Distribution of Items in the Seven Main BIB Blocks
Grade 4

Mode of Comprehension	Type of Text			
	Literary	Informative	Document	TOTAL
Constructs Meaning	19	23	2	44
Extends Meaning	17	5	2	24
TOTAL	36	28	4	68

Table 2-3
Text Type by Mode of Comprehension Distribution of Items in the Seven Main BIB Blocks
Grade 8

Mode of Comprehension	Type of Text			
	Literary	Informative	Document	TOTAL
Constructs Meaning	9	51	5	65
Extends Meaning	7	19	4	30
TOTAL	16	70	9	95

Table 2-4
Text Type by Mode of Comprehension Distribution of Items in the Seven Main BIB Blocks
Grade 12

Mode of Comprehension	Type of Text			
	Literary	Informative	Document	TOTAL
Constructs Meaning	14	58	8	80
Extends Meaning	9	16	5	30
TOTAL	23	74	13	110

2.3.4 Reading Item Development

The development of items for the 1990 assessment began with the identification of critical aspects of text via text mapping procedures. Using the text mapping procedures, item writers diagrammed each passage to depict the relationships among various elements of a text and to clarify distinctions between more and less important text elements.

Using the results of the text mapping procedure, assessment items were written to focus on critical rather than trivial aspects of a text. Close attention was paid to the difficulty of the vocabulary and concepts contained in each item. Although most of the items were presented in multiple-choice format, some of the items were open-ended in response to educators' interest in emphasizing reading-writing connections.

The 1990 assessment included items designed to have readers construct general meanings of varied kinds and levels of difficulty from the text. For informational texts, typical items focused on main ideas, author's purpose and related sequences of details, relations among various parts of the text, support for conclusions or inferences, and relations between problems and their resolution. For stories, typical assessment items focused on themes, important elements in the plot, setting details, character traits and motives, and logical predictions. For documents, readers were asked to use facts, symbols, and spatial relations to perform tasks that simulate real-life situations requiring the use of documents—such as completing a form or determining plans based on a schedule of events.

The 1990 assessment also included items designed to have readers extend and examine the meaning they derived from various kinds of text. For informational texts, typical items focused on noting similarities and contrasts in the information presented within a single text or across multiple sources. For stories, typical items focused on analyzing themes, inferring motives for a character's actions, or identifying the subtle impact of the setting on the reader's interpretation. For documents, typical items asked students to evaluate and make decisions based on the information presented through text and graphic elements.

In structuring item sets, consideration was given to the order of questions following a passage. Constructing meaning questions were typically placed first, asking readers to consider the major themes or purposes of a given passage. Successive questions in the examining or extending meaning modes generally required readers to return to the text to enlarge and enrich these initial understandings. Thus, both the development of items and their organization were patterned after the kinds of comprehension strategies highlighted in the assessment objectives. These guidelines were helpful in assembling the items for the field test, particularly in determining an approximate difficulty level for the items.

Once the field test data had been analyzed, trend items were selected from the 1988 reading assessment and new items were selected that complemented the pool of trend items. Trend items were selected with careful attention to their psychometric characteristics as evaluated in prior assessments. The new items were selected on the basis of the empirical item analyses from the field test as well as the type of stimulus material. One factor that had to be taken into consideration in compiling the items into blocks was that at least one trend and one new block at each age level had to overlap with the age level above and/or below it. There were to be at least three blocks of trend items at each age level as well. In addition to the

criteria for overlap across age levels, trend blocks were selected to represent the best content coverage across the objectives and be representative of the range of performance in previous assessments.

Once the items were selected and the blocks assembled, they underwent final review by ETS subject-area specialists and test editors as well as a review to detect the presence of any bias according to the *ETS Standards for Quality and Fairness*. After internal review, the OMB clearance package was prepared and submitted.

The 1990 reading assessment was based on seven 15-minute blocks of passages and items at each grade level. Fourth graders were asked to respond to 68 items (two of them open-ended) pertaining to 24 brief passages, 11 literary and 13 informational. Eighth graders were asked to respond to 100 items pertaining to 27 brief passages (six literary and 21 informational) and two full-page advertisements. Twelfth graders were asked to read the same advertisements and 26 passages (four literary and 22 informational) and respond to 110 items. At grades 8 and 12, only one of the items was open-ended.

2.3.5 Background and Attitude Questions

Research indicates that school, home, and attitudinal variables affect students' reading comprehension and literacy. Therefore, in addition to assessing how well students read, it is important to understand the instructional context in which reading takes place, students' home support for literacy, and their reading habits and attitudes. To gather contextual information, NAEP assessments include background questions designed to provide insight into the factors that may influence reading proficiency in the literary, informational, and document categories assessed.

Two sets of background questions were included in the 1990 reading assessment. First, every assessed student received a five-minute set of questions requesting basic demographic information as well as other general background information relevant to educational achievement (see section 2.6). Second, each student participating in the reading assessment was asked a five-minute set of background questions pertaining specifically to reading habits, strategies, instruction, and home support. Like the cognitive items, these background questions were submitted to extensive review and field testing. Recognizing the validity problems inherent in self-report data, particular attention was given to developing questions that were meaningful and unambiguous and that would encourage accurate reporting.

Specifically, the reading background questions asked students to report the availability of various reading materials at school—such as newspapers, magazines, textbooks, workbooks, trade books, and study guides—and the extent to which these materials were used in their classes in all subject areas. Students were also asked the extent to which their teachers engaged in various instructional activities before, during, and after reading. Questions about the home reading environment asked students whether they subscribed to magazines at home or had books of their own, were read to when they were young, or read aloud to someone at home. Other questions asked students to report on their attitudes toward reading and the extent to which they read books, magazines, newspapers, and other materials in their leisure time.

2.4 DEVELOPING THE MATHEMATICS ASSESSMENT

2.4.1 Context for Planning the 1990 Mathematics Assessment

In 1988 Congress passed new legislation for NAEP which included—for the first time in the project's history—a provision authorizing voluntary state-by-state assessments on a trial basis in addition to the national assessments that NAEP had conducted since its inception. Anticipating this legislation, in mid-1987 the federal government arranged for a special grant from the National Science Foundation and the Department of Education to the Council of Chief State School Officers (CCSSO) to lay the groundwork for the Trial State Assessment.

The CCSSO established the National Assessment Planning Project to oversee the work for the Trial State Assessment. The National Assessment Planning Project, whose members included policymakers, practitioners, and citizens nominated by 18 national organizations, had two primary purposes. The first was to recommend objectives for the state-level mathematics assessment, and the second was to make suggestions for reporting state results. However, rather than focusing exclusively on the eighth-grade objectives for the Trial State Assessment, the project developed new objectives for all three grades to be assessed in 1990 (fourth, eighth, and twelfth) because the assessment objectives had to be coordinated across all grades. The objectives for the Trial State Assessment Program were the same as for the eighth-grade national program.

2.4.2 Assessment Design Principles

A Mathematics Objectives Committee—comprising a teacher, a school administrator, mathematics education specialists from various states, mathematicians, parents, and citizens—was created by the CCSSO to recommend objectives for the assessment.

Two principles emerged during the discussions of the Mathematics Objectives Committee and became the basis for structuring the objectives and framework for the 1990 assessment. The first principle was that a national assessment, designed to provide state-level comparisons, should not be directed toward measuring only those topics and skills already in the objectives of all states or geared to the *least common denominator* of student preparation. The second principle was that the assessment should also not be used to steer instruction toward one particular pedagogical or philosophical viewpoint to the exclusion of others that are widely held.

The objectives development was also guided by several other considerations: the assessment should reflect many of the states' curricular emphases and objectives; be inclusive of what various scholars, practitioners, and interested citizens believe should be included in the curriculum; and maintain some of the content of prior assessments to permit reporting of trends in performance. Accordingly, the committee gave attention to several frames of reference:

- states' goals and concerns, as reflected through analyses of state mathematics curriculum guides and the recommendations of state mathematics specialists;

- a report on "Issues in the Field," based on telephone interviews with leading mathematics educators, and a draft assessment framework provided by a subcommittee of the Mathematics Objectives Committee;

- the draft of the *Curriculum and Evaluation Standards for School Mathematics*, developed by the National Council of Teachers of Mathematics through intensive work by leading mathematics educators in the United States (National Council of Teachers of Mathematics, 1987); and

- the design of the 1986 mathematics assessment (National Assessment of Educational Progress, 1987). The framework for the 1986 assessment had 35 cells—seven content and five process areas. Because there were 35 cells, the weightings assigned to some of the cells in the 1986 mathematics framework did not result in a sufficient number of items to provide reliable measures of students' knowledge and skills. As a result, it was decided that the matrix guiding the development of the 1990 assessment had to be simplified, and that necessary complexity could be reflected through the designation of specific abilities and topics in each content area.

2.4.3 Assessment Development Process

The framework, objectives, and a set of sample questions developed by the Mathematics Objectives Committee were distributed to the mathematics supervisor in each of the 50 State Education Agencies. These supervisors convened a panel that reviewed the draft objectives and returned comments and suggestions to the project staff. Copies of the draft were also sent to 25 mathematics educators and scholars for review. The Mathematics Objectives Committee incorporated the comments and revisions, and formulated their final recommendations which were approved by the National Assessment Planning Project Steering Committee.

The framework and objectives were then submitted to the National Center for Education Statistics (NCES), which forwarded them for review to the APC. The board approved the objectives with minor provisions about the feasibility of full implementation.[1] The framework and objectives were refined by NAEP's Item Development Panel, reviewed by the Task Force on State Comparisons, and resubmitted to NCES for adoption.

2.4.4 Framework for the Assessment

The framework adopted for the 1990 mathematics assessment is organized according to three mathematical abilities and five content areas. The mathematical abilities assessed are conceptual understanding, procedural knowledge, and problem solving. Content is drawn

[1]This action is contained in a statement issued by the Assessment Policy Committee's Executive Committee on April 29, 1988. The recommendations were ratified by the full committee on June 18, 1988, with two stipulations: that the objectives be so weighted as to permit reporting on trends in performance; and, with regard to the use of calculator-active items and open response questions, that the assessment be developed within the resources available for its administration.

primarily from elementary and secondary school mathematics up to, but not including, calculus. The content areas assessed are numbers and operations; measurement; geometry; data analysis, statistics, and probability; and algebra and functions.

2.4.5 Distribution of Assessment Items

The assignment of percentages of the assessment items that would be devoted to each mathematical ability and content area is an important feature of the assessment design because such weighting reflects the importance or value given to each area at each grade level. For 1990, the National Assessment Planning Project was interested in creating an assessment that would be forward-thinking and could lead instruction; thus, more emphasis was given to problem solving than in previous assessments. Also, participants in the process advised that greater emphasis be given to geometry and to algebra and functions, and less to numbers and operations than in the past.

The distribution of items recommended in the framework by mathematical ability was the same across all grades: Forty percent of the assessment items were to be devoted to conceptual understanding, 30 percent to procedural knowledge, and 30 percent to problem solving. However, the distribution of items by content area varied across grade levels, as shown in Table 2-5.

Table 2-5
Percentage Distribution of Items by Grade and Mathematical Content Area
as Recommended in tne Mathematics Framework

Mathematical Content Area	Grade		
	4	8	12
Numbers and Operations	45%	30%	25%
Measurement	20%	15%	15%
Geometry	15%	20%	20%
Data Analysis, Statistics, and Probability	10%	15%	15%
Algebra and Functions	10%	20%	25%

2.4.6 Developing the Cognitive Items

The 1990 mathematics assessment included open-ended and multiple-choice items. The open-ended items were designed to provide an extended view of students' mathematical knowledge and skills. Building on the recommendations from the CCSSO report, the NAEP Item Development Panel created open-ended items to assess objectives in the framework that could not be measured using multiple-choice items (e.g. ability to draw graphs and figures, generate informal proofs or generalize relationships). In addition, some open-ended items were designed to provide insight into the ways in which students think about mathematics (e.g.,

students were asked to describe the procedures they used to arrive at answers to selected problems.) Although time-consuming to analyze, these descriptions provide an understanding of the ways in which students reach correct and incorrect answers.

At each grade level, the 1990 mathematics assessment included 10 different 15-minute segments or "blocks" of multiple-choice and open-ended items. Two of the 10 blocks were designed to be answered using a calculator and three were accompanied by a paced audiotape to assess students' estimation skills and provide for complex problem-solving situations. Because the blocks contained a variety of item types, there were no rigid criteria dictating parallel structure across blocks. All of the blocks, except for the three blocks accompanied by a paced audiotape, were assembled three to a booklet and each student was asked to respond to one booklet. These seven blocks (including the two requiring calculators) are balanced across seven booklets (for the main BIB design). These seven booklets at grade 8 also were used for the Trial State Assessment. The three blocks accompanied by audiotape were assembled into one booklet with a fourth block which was one of the other seven blocks.

Items appearing in blocks that allowed the students to use calculators attempted to assess not only the ability to use a calculator correctly but also the ability to choose the appropriate procedure, i.e., to decide which is the most appropriate method for solving a particular problem—calculator, paper and pencil, or mental arithmetic and estimation.

The 1990 mathematics assessment contained 143 questions in total at grade 4, 41 of which required open-ended responses (109 items in total with 28 open-ended ones for the seven main BIB blocks). At grade 8, there were 191 questions, of which 42 were open-ended (137 items in total including 35 open-ended items for the seven main BIB blocks). Twelfth graders were administered a total of 203 questions, of which 48 were open-ended (144 in total with 35 open-ended items in the seven main BIB blocks). The Trial State Assessment Program was based on the 137 items in the seven main BIB blocks, including the two requiring scientific calculators, but not the three in the special audio paced-tape study booklet requiring tape recorders.

Tables 2-6, 2-7, and 2-8 provide a description of the number of items for each content and ability group that was included at each grade level for the main BIB design.

2.4.7 Background Questions

In addition to the cognitive questions, the 1990 mathematics assessment included two 5-minute sets of general and mathematics background questions designed to gather contextual information about students, their experiences in mathematics, and their attitudes toward the subject. Some of the background questions were newly developed for 1990. Others were repeated from previous assessments to permit an analysis of trends across time in students' exposure to mathematics, instructional experiences, and attitudes toward the subject.

Every student booklet for the mathematics assessment began with a common core of questions which asked general background questions (see section 2.6). Another set of questions,

Table 2-6
Content-by-ability Distribution of Items in the Seven Main BIB Blocks
Grade 4

Mathematical Ability	Content Areas					
	Numbers and Operations	Measurement	Geometry	Data Analysis, Statistics, and Probability	Algebra and Functions	TOTAL
Conceptual Understanding	18	6	8	4	6	42
Procedural Knowledge	16	9	1	3	2	31
Problem Solving	18	6	5	1	6	36
TOTAL	52	21	14	8	14	109

Table 2-7
Content-by-ability Distribution of Items in the Seven Main BIB Blocks
Grade 8

Mathematical Ability	Content Areas					
	Numbers and Operations	Measurement	Geometry	Data Analysis, Statistics, and Probability	Algebra and Functions	TOTAL
Conceptual Understanding	18	7	13	9	12	59
Procedural Knowledge	15	9	4	5	8	41
Problem Solving	12	5	9	5	6	37
TOTAL	45	21	26	19	26	137

Table 2-8
Content-by-ability Distribution of Items in the Seven Main BIB Blocks
Grade 12

Mathematical Ability	Content Areas					
	Numbers and Operations	Measurement	Geometry	Data Analysis, Statistics, and Probability	Algebra and Functions	TOTAL
Conceptual Understanding	12	7	11	9	14	53
Procedural Knowledge	14	9	5	7	13	48
Problem Solving	11	7	8	6	11	43
TOTAL	37	23	24	22	38	144

grouped into three categories, were included in the second five-minute section of mathematics background questions:

> *Time Spent Studying Mathematics*: Time spent on task and mathematics coursework has been shown to be strongly related to mathematics achievement (Raizen & Jones, 1985). Students were asked to describe both the amount of instruction they receive in mathematics and the time spent on mathematics homework.
>
> *Instructional Practices*: The nature of students' mathematics instruction is also postulated to be related to achievement (Dossey, Mullis, Lindquist, & Chambers, 1988). Students were asked to report their experience in using various instructional materials in the mathematics classroom, including calculators, models, and manipulatives. In addition, they were asked about the instructional practices of their mathematics teachers and the extent the students themselves practice the communication of mathematical ideas—such as writing out explanations, justifications, or proofs—in their mathematics classes.
>
> *Attitudes Towards Mathematics*: Students' enjoyment of and confidence in mathematics and their perceptions of the usefulness of the discipline to their present and future lives appear to be related to mathematics achievement (Tobias, 1987). Students were asked a series of questions about their attitudes and perceptions about mathematics, such as whether they enjoy mathematics and whether they are good in mathematics.

2.5 DEVELOPING THE SCIENCE ASSESSMENT

2.5.1 Science Objectives

NAEP's objectives for the 1990 science assessment were the result of a consensus process involving many individuals, ranging from science educators and specialists to members of the lay public. An Assessment Development Panel began the objectives development process by revising the framework from the 1986 assessment and preparing an initial set of science objectives. The draft objectives were then reviewed by the Item Development Panel and by external consultants from across the country, representing various interests and concerns. The Item Development Panel, a committee of science educators, administrators, and researchers, was established to develop items for the 1990 assessment using the framework provided in these objectives. Among the reviewers for the 1990 objectives were teachers, school science coordinators, scientists, school administrators, state science supervisors, and members of the APC. The Science Assessment Development Panel remained involved throughout the review and revision process.

The 1990 NAEP science assessment objectives reflected the consensus among educators that school science should help students acquire the knowledge, skills, and understandings necessary to fulfill their human, social, and economic responsibilities. The objectives were based on an interactive model in which scientific habits of mind serve as a filter between the features of the learner—including his or her cognitive abilities, attitudes, and home and school

experiences—and the outcomes of science learning. Three elements of the model were included in the assessment—science knowledge, scientific habits of mind, and the ability to solve problems and conduct inquiries.

The first element, **Science Knowledge**, emphasizes that students' knowledge base in science should include information not only about the natural phenomena that are the objects of study in the major scientific disciplines—life sciences, physical sciences, and earth and space sciences—but also the fundamental concepts, principles, and theories in these disciplines. Second, students' knowledge base should contain information about the nature of science, including a recognition of the characteristics of science that set it apart from other human activities—in particular, its empirical and theoretical methods and its philosophy. Finally, students should be informed about the history of science and the interactions among science, culture, society, and technology.

The second element, **Scientific Habits of Mind**, includes both the ability to think scientifically and the inclination to do so beyond the confines of the science classroom. Inductive and deductive reasoning; verbal, analogical, and spatial reasoning; and creative thinking are among the primary elements of scientific thinking.

The third element, **Solving Problems and Conducting Inquiries**, encompasses a wide range of activities, from the novice efforts of students interacting with the natural world to the work of experienced scientists. Although these thinking skills are often exercised as a part of everyday experience, for the purpose of the assessment, discussions of students' ability to solve problems and conduct inquiries were limited to academic applications. The capacity to wield these skills in various contexts is considered by science educators to be an important indicator of science achievement.

Within the context of these three elements, the Assessment Development Panel established a framework for the assessment of student performance based on the usual structuring of the science curriculum (i.e., according to traditional disciplines) and the anticipated utility of the findings for policymakers and science educators. The framework was a two-dimensional matrix plotting three content areas (Life Sciences, Physical Sciences, and Earth and Space Sciences) by three thinking skills (conducting inquiries, solving problems, and knowing science). The Nature of Science was considered to be a category that overlapped all content areas and thinking skills. To guide the development of assessment items, the panel assigned weights to each of the major categories in the framework, reflecting the relative importance of each of the content and cognitive areas assessed.

2.5.2 Science Materials

The 1990 science assessment included both multiple-choice questions and several types of open-ended questions. Many multiple-choice questions required students to respond to stimulus material such as figures, graphs, and data tables. Some open-ended questions required short responses, two required longer essay responses, and others were figural response questions in which students were required to draw arrows on lines on figures and graphs in response to particular tasks.

2.5.3 Science Item Development

The development of the 1990 NAEP science assessment reflected the consensus view of the Assessment Development Panel and Item Development Panel that any assessment of students' science knowledge can cover only a small part of the domain of science education provided across the school years. Therefore, the panel prioritized the content and cognitive areas included in the assessment. Weighting decisions were based on information about the typical science curriculum at each grade level; on NAEP's legislatively mandated obligation to provide information on trends in science achievement; on measurement requirements for reporting content area subscales; and on the views of the Assessment Development Panel, the Item Development Panel, and NAEP staff and consultants. Weights were assigned separately at each grade level because the content emphases and cognitive expectations appropriate for fourth-, eighth-, and twelfth-grade students are different. The cognitive classification of each assessment item was based on the judgments of NAEP staff and consultants about the thinking skills that the average student in the target population would be likely to use when answering the item.

The distribution of items recommended in the framework by science thinking skills was the same across all grades: Forty percent of the assessment was to be devoted to knowing science, 40 percent to solving problems, and 20 percent to conducting inquiries. However, the distribution of items by content area varied across grade levels, as shown in Table 2-9.

Table 2-9
Percentage Distribution of Items by Grade and Science Content Area
as Recommended in the Science Framework

Category	Grade		
	4	8	12
Nature of Science	10%	10%	12%
Life Sciences	30%	30%	32%
Physical Sciences	30%	30%	34%
Earth and Space Sciences	30%	30%	22%

Based on the weightings established, the Assessment Development Panel, Item Development Panel, and additional consultants began to write items for the assessment. All new items were subsequently reviewed by ETS subject-area and measurement specialists, sensitivity reviewers, and test editors, thoroughly field-tested, and revised as needed subsequent to each of these stages of review.

Because the assessment cannot include items on all appropriate topics, the final selection of assessment questions was guided by the Assessment Development Panel and the Item

Development Panel. The selection was based on the careful balancing of a number of requirements, including:

- the need to repeat items from previous assessments to enable reporting on trends in science achievement across time;
- the need to meet distribution specifications set by the matrix dimensions of content and thinking skills, and to integrate the other priorities defined by the panel; that is, the inclusion of items that pertain to the history of science, context, technology, and mathematical content and abilities; and
- the need to offer items that reflect a wide range of difficulty.

The final choice of questions reflected the professional judgments of NAEP staff and consultants as to the quality of individual items and the centrality of the thinking skills and content that each measures. The final pool of questions at fourth grade included 95 multiple-choice items, 7 open-ended items, and 10 figural-response items. The final pool of questions at eighth grade included 122 multiple-choice items, 10 open-ended items, and 14 figural-response items. At twelfth grade, the final pool of questions included 126 multiple-choice items, 8 open-ended items, and 16 figural-response items.

After the final selection of assessment items, the items were assembled into seven 15-minute blocks at each grade level. One of the seven blocks at each grade was composed exclusively of figural-response items. Each block was again reviewed by ETS subject-area and measurement specialists, a sensitivity reviewer, and a test editor, according to the *ETS Standards for Quality and Fairness*. Tables 2-10, 2-11, and 2-12 provide a description of the number of items for each content and skill group that was included at each grade level for the main BIB design.

2.5.4 Science Background and Attitude Questions

The 1990 science assessment included a questionnaire at each grade level that provided information about the amount of time that students spent studying science, the instructional practices used in their science classes, the types of laboratory activities they participated in, their attitudes toward science learning, and their exposure to informal science learning. Decisions about the specific areas to be assessed were made with guidance from the Science Assessment Panel and the Item Development Panel.

2.6 DEVELOPING THE COMMON CORE STUDENT QUESTIONS AND QUESTIONNAIRES

In 1990, NAEP administered a series of questionnaires to students, teachers, and principals or other administrators. Similar to the development of the cognitive items, the development of the policy issues and items was an iterative process that involved staff work, field testing, and review by external advisory groups. A Policy Analysis and Use Panel drafted a

Table 2-10
Content-by-skill Distribution of Items in the Seven Main BIB Science Blocks
Grade 4

Thinking Skills	Content Areas			
	Life Sciences	Physical Sciences	Earth and Space Sciences	Nature of Science
Knowing Science	10	9	18	10
Solving Problems	14	20	12	0
Conducting Inquiries	6	8	3	2
TOTAL	30	37	33	12

Table 2-11
Content-by-skill Distribution of Items in the Seven Main BIB Science Blocks
Grade 8

Thinking Skills	Content Areas			
	Life Sciences	Physical Sciences	Earth and Space Sciences	Nature of Science
Knowing Science	10	10	19	17
Solving Problems	20	16	20	0
Conducting Inquiries	14	15	4	1
TOTAL	44	41	43	18

Table 2-12
Content-by-skill Distribution of Items in the Seven Main BIB Science Blocks
Grade 12

Thinking Skills	Content Areas			
	Life Sciences	Physical Sciences	Earth and Space Sciences	Nature of Science
Knowing Science	9	16	13	13
Solving Problems	18	27	21	3
Conducting Inquiries	15	9	5	1
TOTAL	42	52	39	17

set of policy issues and made recommendations regarding the design of the questions. They were particularly interested in capitalizing on the unique properties of NAEP and not duplicating other surveys (e.g., The School and Staffing Surveys). For the 1990 assessment, the resulting framework focused on six educational areas: curriculum, instructional practices, teacher qualifications, educational standards and reform, school conditions, and conditions outside of the school that facilitate learning and instruction.

The questionnaires developed were as follows:

The **student demographics (common core) questionnaire** (18 questions at grade 4, 21 questions at grade 8, and 30 questions at grade 12) included questions about race/ethnicity, language spoken in the home, mother's and father's level of education, reading materials in the home, homework, attendance, school climate, academic expectations, which parents live at home, and which parents work. In many cases the questions used in 1990 were continued from prior assessments. Although many of the questions were common to the three age/grade levels assessed, some were specifically targeted to a particular level.

The **student content area questionnaire** (ranged from 14 to 35 questions, depending on the grade level and subject area) included questions about instructional activities, courses taken, use of specialized resources such as calculators in mathematics class, and views about utility and value of the subject matter. These questionnaires were specific to each subject area and have been described in the preceding sections.

To promote efficiency in data collection and to enable linking student background and achievement data, the two student questionnaires were included in the assessment booklets and administered in the same session as the cognitive items. With the exception of fourth graders, students were given five minutes to respond to each questionnaire. To improve the validity of fourth graders' answers to the questionnaire about their demographic backgrounds, the assessment administrators read the questions aloud and clarified the vocabulary and intent of the questions. At the other two age/grade levels, only the first race/ethnicity question was read aloud; students read and answered the remaining questions on their own during the five-minute time period.

To supplement the information on instruction as reported by students, the teachers of fourth and eighth graders participating in the mathematics assessments (including the Trial State Assessment at grade 8) were asked to complete a teacher questionnaire about their instructional practices, teaching backgrounds, and characteristics. In addition, administration of an eighth-grade science teacher questionnaire was made possible by supplemental funding from the National Science Foundation through a subcontract from Horizon Research, Inc. to ETS.

The teacher questionnaires contained two parts. The first part pertained to the teachers' background and training. The second part pertained to the procedures used for *each class* containing an assessed student.

The **Teacher Questionnaire, Part I: Background and Training** (at grades 4 and 8, 34 questions for mathematics; at grade 8, 100 questions for science)

included questions pertaining to gender, race/ethnicity, years of teaching experience, certification, degrees, major and minor, coursework in education, coursework in subject area, in-service training, extent of control over classroom, instruction, and curriculum, and availability of resources for classroom.

The **Teacher Questionnaire, Part II: Classroom by Classroom Information** (at grade 4, 34 questions; at grade 8, 35 questions for mathematics and 58 questions for science) included questions on the ability level of students in the class, whether students were assigned to the class by ability level, time on task, homework assignments, frequency of instructional activities used in class, instructional emphasis given to the topics and skills covered in the assessment, and use of particular resources.

A school characteristics and policy questionnaire was given to the principal or other administrator of each school that participated in the 1990 NAEP assessment. Including the schools in the Trial State Assessment, 6,000 school questionnaires were administered in 1990.

The **school characteristics and policies questionnaire** (at grades 4 and 8, 117 questions and at grade 12, 125 questions) included questions about background and characteristics of school principals, length of school day and year, school enrollment, absenteeism, drop-out rates, size and composition of teaching staff, policies about tracking, curriculum, testing practices, special priorities and school-wide programs, availability of resources, special services, community services, policies for parental involvement, and school-wide problems.

For each student excluded from the assessment, including those in the Trial State Assessment Program, schools were required to complete a questionnaire about the characteristics of that student and the reason for exclusion.

The **excluded student questionnaire** collected data about students' race/ethnicity and the reason for exclusion from the assessment. For students with an Individualized Education Plan (IEP), the questionnaire included questions about students' functional grade level, mainstreaming, and special education programs. For Limited English Proficient (LEP) students, it asked about students' native language, time spent in special education and language programs, and the level of the students' English language proficiency.

2.7 FINAL PREPARATION OF THE 1990 ASSESSMENT MATERIALS

The field test for the 1990 program was conducted in January through March, 1989. Generally, field tests were conducted at all three levels in each school district that had two classes at fourth grade and one class at each of grades 8 and 12. Field tests were carried out in 104 fourth-grade classes, 88 eighth-grade classes, and 80 twelfth-grade classes. Approximately 300 responses were obtained to each question in the field test materials.

The data were collected, scored, and analyzed in preparation for meetings with the respective Panels. Based on the empirical item analysis, which provides the mean percentage of

correct responses for each item in the field test, committee members, test development, and NAEP staff reviewed the materials with four purposes: to determine which items were most related to achievement in the three subject areas; to determine the need for revisions of items that lacked clarity, or had ineffective item formats; to prioritize items to be included in the 1990 assessment; and to determine appropriate timing for assessment items.

Once the committees had selected the items, all questions were rechecked for content, measurement, and sensitivity concerns. The OMB clearance process was initiated in June 1989 with the submission of draft materials to NCES. The final package was submitted in July 1989. Throughout the clearance process, revisions were made in accordance with changes required by the government.

Chapter 3

SAMPLE DESIGN

Keith F. Rust

Westat, Inc.

The samples for the 1990 NAEP assessment were selected using a complex multistage sample design involving the sampling of students from selected schools within 97 selected geographic areas, called primary sampling units (PSUs), across the United States.

The sample design had four stages of selection:

1) selection of geographic PSUs (counties or groups of counties);

2) selection of schools within PSUs;

3) assignment of session types to schools; and

4) selection of students for session types within schools.

The samples were drawn for the three different age classes, and for each age class the samples were of two distinct types. The first type consisted of the cross-sectional or "main" samples, while the second type consisted of the trend or "bridge" samples. The populations surveyed with each of these sample types are defined in Table 1-1. Separate samples of schools were required for the bridge samples and main samples, because of various differences in the calendar period for test administration, the format of the administration, and, in the case of age class 17, the grade and age definition of the population of interest.

In addition to representing the respective populations as a whole, for the main samples there was oversampling of private schools, and public schools with moderate or high enrollment of Black or Hispanic students (see section 3.2). This oversampling was undertaken to increase the sample sizes of private school students and minority students, so as to increase the reliability of estimates for these groups of students. There was a further subpopulation for which it was desired to ensure adequate representation in the sample. This was the population of students in those 38 states that participated in the Trial State Assessment Program in 1990. This population was of interest because of the requirement to equate the samples of eighth-grade students taking the mathematics assessment, across the two components, national and state-by-state. The aim of ensuring adequate representation was achieved by augmenting the sample of PSUs. The procedure for doing this is discussed in section 3.1.

The overall assessment period fell into three time periods—fall, winter, and spring. Not all assessment components were conducted in each time period. Table 3-1 shows the relationship between the various sample components and the assessment periods. The sizes of the PSU and school samples and the procedures for their selection were determined by the assessment period, as well as by the population to be surveyed and the method of administration in each case.

Table 3-1
Assessment Type by Age Class and Assessment Period

Age Class	Fall 10/9/89 - 12/15/89	Winter 1/8/90 - 3/16/90	Spring 3/19/90 - 5/18/90
9	—	Main assessment (half) Bridges to 1984, 1986	Main assessment (half)
13	Bridges to 1984, 1986	Main assessment (half)	Main assessment (half)
17	—	Main assessment (half)	Main assessment (half) Bridges to 1984, 1986

The age class 9 and age class 13 bridge samples used the same school and student eligibility requirements as the respective main samples. Nevertheless, special bridge samples were required because:

1) The conditions for administration of the assessment varied considerably between the main sample and bridge sample sessions.

2) The need in the bridge samples for six distinct session types for each of these two age classes, together with the need for three distinct session types for the main samples for each age class, made it infeasible to conduct both main sample sessions and bridge sessions in a given school.

3) For age class 13, the main samples were conducted at an inappropriate time of the year for bridging purposes, so that a distinct sample of schools was needed to undertake the bridge assessments in the fall of 1989.

A separate sample of schools was required for the bridge sessions and the main sessions for age class 17 primarily because the definitions for student eligibility, based on age and grade, differed substantially between the two samples, even though the same population of schools was surveyed in each case. Conditions of administration also varied somewhat, and there were five distinct bridge session types and three distinct main sample session types. Thus, it was not feasible to conduct main sample and bridge sample sessions within a single school.

This chapter gives details of the sample selection procedure, and information on the results of the sampling process. Fuller details are given in the report *1990 National Assessment of Educational Progress Sampling and Weighting Procedures, Part 2: National Assessment, Final Report* (Rust, Burke, & Fahimi, 1992).

3.1 PRIMARY SAMPLING UNITS

In the first stage of sampling, the United States (the 50 states and the District of Columbia) was divided into geographic primary sampling units (PSUs). Each PSU met a minimum size requirement (a population of at least 60,000 in the 1980 Census) and comprised either a metropolitan statistical area (MSA), a single county, or (more usually in the case of nonMSA PSUs) a group of contiguous counties. In the case of New England MSAs, which are not formed from whole counties, the corresponding New England County Metropolitan Areas, which are defined in terms of whole counties, were designated as the PSUs. The New York City MSA was divided along county/borough lines into three PSUs for reasons of administrative and sampling convenience. Each PSU was contained entirely within one of the four regions defined in Table 3-2. These regions were used to stratify the sample of PSUs, ensuring that each region was adequately represented in the various assessment samples.

Table 3-2
Geographic Regions Used for Stratification

Northeast	Southeast	Central	West
Connecticut	Alabama	Illinois	Alaska
Delaware	Arkansas	Indiana	Arizona
District of Columbia	Florida	Iowa	California
Maine	Georgia	Kansas	Colorado
Maryland	Kentucky	Michigan	Hawaii
Massachusetts	Louisiana	Minnesota	Idaho
New Hampshire	Mississippi	Missouri	Montana
New Jersey	North Carolina	Nebraska	Nevada
New York	South Carolina	North Dakota	New Mexico
Pennsylvania	Tennessee	Ohio	Oklahoma
Rhode Island	Virginia*	South Dakota	Oregon
Vermont	West Virginia	Wisconsin	Texas
Virginia*			Utah
			Washington
			Wyoming

* That part of Virginia that is part of the Washington, DC-MD-VA metropolitan statistical area is included in the Northeast region; the remainder of the state is included in the Southeast region.

In a few cases an MSA crossed region boundaries. Such MSAs were split into two or more PSUs as necessary (e.g., the Cincinnati OH-KY-IN MSA was split into the Cincinnati OH-IN PSU in Region 3 and the Cincinnati KY PSU in Region 2). Twelve subuniverses of PSUs were then defined as described below.

The 28 largest PSUs were included in the sample with certainty. An additional six very large PSUs (four from the Southeast and two from the West) that had large proportions of Black students and/or Hispanic students were also included with certainty. The 34 certainty PSUs constituted 32 metropolitan areas, since the New York City MSA was divided into three certainty PSUs. The inclusion of these 34 PSUs in the sample with certainty provided an approximately optimum cost-efficient sample of schools and students when samples were drawn

within them at the required national sampling rate. The representativeness of the sample for minority groups was enhanced by ensuring that these PSUs were included in the sample, since these minority groups are relatively heavily represented within these certainty PSUs. The remaining smaller PSUs were not guaranteed to be selected for the sample. These were grouped into a number of noncertainty strata (so called because the PSUs in these strata were not included in the sample with certainty), and sample PSUs were selected from each stratum.

The PSUs were classified into four regions, each containing about one-fourth of the U.S. population. These regions were defined primarily by state (Table 3-2). In each region, noncertainty PSUs were classified as MSA or nonMSA. In the Southeast and West regions, the PSUs in which the combined proportion of population which were Black and Hispanic respectively in the 1980 Census exceeded 20 percent, were classified as high minority. The resulting major strata, or subuniverses, are shown in Table 3-3.

Table 3-3
The Sampling Subuniverses
and the Number of Noncertainty Strata in Each

Region	MSA PSUs		NonMSA PSUs	
	Regular Strata	High-minority Strata	Regular Strata	High-minority Strata
Northeast	8	—	2	—
Southeast	4	6	4	6
Central	8	—	6	—
West	4	6	4	2
Total	24	12	16	8

Within each major stratum (subuniverse), further stratification was achieved by ordering the noncertainty PSUs according to several additional socioeconomic characteristics, yielding 60 strata. The number of such strata formed within each subuniverse is shown in Table 3-3. The strata were defined so that the aggregate of the measures of size of the PSUs in a stratum was approximately equal for each stratum, except for strata in the high-minority subuniverses, in which the aggregate was approximately half that of the regular strata. The size measure used was the population from the 1980 Census. The characteristics used to define strata were the percent minority population, the percentage change in total population since 1970, the per capita educational expenditure, and the percent of persons employed in manufacturing (MSA subuniverses only) and the percentages of rural and urban dwellers (nonMSA subuniverses only). A total of 63 PSUs was sampled from the 60 noncertainty strata. Initially, one PSU was selected with probability proportional to size from each of the 60 noncertainty strata. That is, within each stratum, a PSU's probability of being the sample selection from that stratum was proportional to its population. Thus the high-minority subuniverses were sampled at approximately twice the rate of the other subuniverses, since they were about half as large. This procedure of oversampling from the high-minority subuniverses was used with the aim of reducing somewhat the level of sampling error for estimates relating to the populations of Black and Hispanic students.

Three additional PSUs were selected, with the aim of decreasing the variance of estimates made from the national NAEP samples, for the aggregate of 39 states (including the District of Columbia) that were expected to participate in the 1990 Trial State Assessment. Although the subset of 48 sampled PSUs that fell within this set of states constituted an unbiased sample, through random chance certain population segments within the 39 states were not well represented by these 48 PSUs. A supplementary sampling scheme was developed, using the subuniverse as the unit level at which we wished to ensure good PSU representation. The rule for supplementation within each subuniverse was as follows: The minimum number of supplementary PSUs was chosen so that, after supplementation was implemented, the number of PSUs from participating states was at least 75 percent of the expected number of PSUs from participating states in the original sample.

Applying this rule, no supplementary PSUs were required in 10 of the 12 subuniverses. In the nonMSA Northeast subuniverse, neither of the two originally selected PSUs was from a state expected to participate in the Trial State Assessment, whereas the expecteo number was 0.97. Thus, one supplementary PSU was selected for this subuniverse. In the regular nonMSA Southeast subuniverse, only one of four originally selected PSUs was from a participating state, compared with the expected number of 3.33. Two supplementary PSUs were added as a result.

In each subuniverse, the supplementary PSUs were chosen from among the participating states, using probability proportional to size selection. In the case of the subuniverse for which two supplementary PSUs were chosen, systematic sampling was implemented so as to ensure that the two selections were from different strata and different states. In this case, sampling was restricted to those three strata in the subuniverse for which the initially selected PSU was not from among the states expected to participate. Although supplementation was used in only two subuniverses, in most subuniverses supplementation could potentially have been used had the initial sample of PSUs fallen in a different set of states. It was possible to determine the unconditional probability of selection for each of the 63 PSUs in the sample—that is, the probability that the PSU could have come into the sample, either through the initial selection or through supplementation. These unconditional probabilities were used in weighting the student samples (see Chapter 10). It should be noted that in the final event two states that were expected to participate in the Trial State Assessment dropped out, while one state joined the program at a late date. This had no adverse effect on the PSU representation of the subset of actual participating states; 51 noncertainty PSUs were planned from the population of the 39 states expected to participate, and there were in fact 51 noncertainty PSUs from the population of 38 states that did participate.

The final sample of 97 PSUs was drawn from a population of about 1,000 PSUs. Primarily because of the use of MSAs as PSUs, PSUs varied considerably as to their probability of selection, since they varied greatly in size. The 34 certainty PSUs consisted of the 26 largest MSAs in the country, based on the 1980 population from the Census, plus six other large MSAs from the Southeast and West regions with in excess of 20 percent of their populations being Black or Hispanic. The 36 selected noncertainty MSA PSUs had probabilities of selection ranging from 0.041 to 0.493, while the 27 selected nonMSA PSUs had probabilities ranging from 0.021 to 0.077. The variations in probability depended upon the size of the PSU (1980 population), whether or not the PSU was in a high minority subuniverse, and whether or not the PSU was wholly (or partly) contained within one or more of those states that participated in the Trial State Assessment. Parts of 38 states were included in the 97 selected PSUs.

Samples of 94 PSUs each were drawn at one time for the 1986, 1988, 1990, and 1992 assessments. They were drawn so as to provide for the rotation of the PSUs from one assessment to the next, except that certainty PSUs were retained in each assessment year, and some of the larger noncertainty PSUs were retained for two successive assessment years. The additional three supplementary PSUs for the 1990 sample were chosen subsequently, just prior to the selection of the school sample for 1990. All 97 PSUs were used for both the main assessments and bridge assessments for all three age classes.

3.2 SCHOOLS FOR MAIN SAMPLES; ASSIGNMENT OF SESSIONS TO SCHOOLS

In the second stage of sampling, the public schools (including Bureau of Indian Affairs schools and Department of Defense schools) and private schools (including Catholic schools) within each of the 97 PSUs were listed according to the grade ranges associated with the three age classes. Table 3-4 shows the numbers of schools included in the various frame components. The population of eligible public schools for each age class was the same for bridge and main samples in each case. However, for private schools additional schools were identified and included on the frame for the main samples but not the bridge samples. Any school having one or more of the eligible grades, and located within an appropriate PSU, was included on the frame of schools (the list of schools from which the samples of schools were drawn) for a given sample. For each age class, only a fraction of one percent of age-eligible students was enrolled in ineligible schools. Each school within the 97 PSUs with a grade in the range of 2 to 12 was included in at least one age class—a total of 38,456 schools. An independent sample of schools was selected for each of the age classes. Thus some schools were selected for assessment of two age classes, and a few were selected for all three.

Table 3-4
Grade Definition of School Eligibility for Frame Inclusion
and Frame Sizes, Main and Bridge Samples

Age Class	Frame Included Schools with Any Grade in This Range	Number of Schools on Frame		
		Public*	Private **	
			Bridge	Main
9	2 - 5	18,570	8,671	10,078
13	6 - 9	18,791	9,331	10,694
17	9 - 12	5,758	3,074	4,176
Total	2 - 12	26,960	10,016	11,496

* Public, Bureau of Indian Affairs, and Department of Defense Schools
** Catholic and other private schools

The lists of schools were obtained from several sources. The list of regular public schools was obtained from the National Center for Education Statistics' Common Core of Data (CCD) for 1988. To this were added Bureau of Indian Affairs, Department of Defense,

Catholic, and other private schools from the 1988 list of schools maintained by Quality Education Data, Inc. (QED). The CCD file contained more data on schools which could be used in sampling (specifically, data on minority enrollment and enrollment by grade) than the QED file, but contained only regular public schools.

For the main samples, supplementary lists of private schools were obtained from two sources, and added to the QED list of private schools. This supplementation was undertaken because previous studies have revealed that the QED list of private schools is somewhat deficient in its coverage of non-Catholic private schools (Westat, Inc., 1984, Chapter 4). Although the percentage of students in schools not covered by QED is small as a percentage of the total student population (believed to be less than 1 percent), we believed that it was important to attempt to reduce this noncoverage for the main samples, since separate estimates were to be produced for private schools, based on samples of private school students obtained using a relatively high rate of sampling, compared to that of public school students for the main sample and private school students in the bridge samples and in previous years.

The first supplementary private school listing source used was a list of private schools developed for the National Center for Education Statistics' 1988 School and Staffing Survey. This list was restricted to a sample of counties selected for the survey. Certain of these counties, generally large in population, were also included, independently by chance, in the NAEP sample PSUs. The schools from such counties were added to the NAEP frame after steps were taken to eliminate duplicates with the QED list of private schools. The second source was a series of lists generated clerically from the yellow pages of telephone directories from metropolitan areas included in the NAEP PSU sample. This list was matched against the other sources of private schools to eliminate duplicates.

The process of private school list supplementation resulted in the addition of 1,480 schools to the frame, 693 obtained from the School and Staffing Survey list and 787 derived from telephone directories. The nature of these listing sources meant that little was known about these schools, and in particular the grade span was seldom known. This meant that a large proportion of these schools were in fact out of scope for a given age class school frame. These 1,480 schools were included in the sampling process, and the 113 that were selected in the samples were initially screened via a telephone inquiry to establish whether or not the school was in scope. Fifty-eight schools were dropped from the sample as a result of this process, and a further 22 were subsequently found to be out of scope by the Westat field supervisor who contacted the school regarding participation in NAEP. Thus it appears that only about 30 percent of the supplementary private schools were in scope across all three age class frames, bearing in mind that many such schools were added to all three frames because we had no prior knowledge as to their grade spans.

Schools within each PSU were selected (without replacement) with probabilities proportional to assigned measures of size. Equal measures of size were assigned to schools containing estimates of age/grade-eligible students ranging from 20 to 150 (for age class 9), or to 200 (for age class 13 and age class 17). Schools larger than the indicated maximum size were selected within the selected PSUs with probabilities proportional to the number of age/grade-eligible students. Schools with fewer than 20 estimated age/grade eligibles were assigned somewhat lower measures of size, and thus lower probabilities of selection, since assessment in these schools involved substantially higher per-student administrative costs.

Each public school with minority (Black and/or Hispanic) enrollment in excess of 15 percent of total enrollment was given double the probability of selection of a public school of similar size in the same PSU with minority enrollment below 15 percent. Overall probabilities of selection for such high-minority schools were twice those for other schools of the same size from a given PSU in order to enlarge the sample of Black and Hispanic students, thereby enhancing the reliability of estimates for these groups. For a given overall size of sample, this procedure reduces somewhat the reliability of estimates for all students as a whole and for those not Black or Hispanic.

Each private school was given triple the probability of selection of a low-minority public school of similar size. These greater probabilities of selection were used to ensure adequate samples of private school students in order to allow the derivation of reliable estimates for such students.

The total number of schools selected for each age class was determined to be such that the predesignated student sample sizes would be achieved by selecting all eligible students in a selected school, up to a maximum of 150 (for age class 9) or 200 (for age class 13 and age class 17), allowing for losses due to nonparticipation of selected schools and students and the exclusion of students from the assessment.

In each of the 97 PSUs for each age class, a minimum of three or four schools was selected. These minima were based primarily on the total number of students and hence schools required for the particular age class. The choice between three and four schools as a minimum was determined by the relative sampling rate of schools, from the PSU and age class, in the 1988 National Educational Longitudinal Study (NELS) First Follow-up sample. Schools included in this 1990 NELS sample, when selected by the NAEP sampling algorithm, were dropped from the NAEP sample. Schools were sorted systematically, within PSU and age class, by whether or not they were included in the NELS sample, so that it was possible to determine what proportion of the NAEP sample would be NELS sample schools. The minimum sample size per PSU was set so that, after dropping the NELS schools from the NAEP sample, there would remain at least two and almost certainly three schools. Since the NELS sample schools all contained grade 10, the minimum sample size was almost always set at three for age class 9, and often set at four for age class 17. The process of sampling NELS schools and then dropping them was undertaken to permit the calculation of nonresponse adjustments in the sample weighting to account for the nonoverlap of the NAEP and NELS sample. In fact this procedure was not ultimately used, but rather a set of NELS selection probabilities were derived for the schools selected for NAEP, and these were used instead in the weighting process (see Chapter 10).

The use of these minima for the sample size of schools per PSU was derived as a compromise between two desirable but conflicting objectives. The first of these was to ensure substantial representation from within each selected PSU (after the impact of nonresponse). The second objective was the need to keep the variability in overall student sampling probabilities (and hence weights) to a low level, so as to control the sampling errors associated with NAEP estimates. The use of a large minimum sample size requirement per PSU would act to reduce unduly the weights of students selected from small PSUs.

This design, with the important exceptions described above, had the goal of yielding a sample of students in a given age or grade with approximately uniform probabilities of selection. The efforts to oversample certain subgroups in the population, the practical constraints on the sample size within each school, and the need to ensure an adequate sample within each PSU resulted in some substantial violations of this general goal. The distributions of selection probabilities of the selected students, as reflected in their sampling weights, is discussed in Chapter 10.

For all three age classes, a sample of schools was first drawn for the bridge assessments (see below). These schools were then excluded from the frame when the samples of schools were drawn for the main assessments. Adjustments were made to the sampling weights to reflect the appropriate probabilities of selection to yield unbiased estimates for both bridge and main samples. Schools selected for main assessments were further classified as belonging to the winter main assessment or the spring main assessment. Random systematic half-samples of the schools in each PSU were assigned respectively to the winter and spring assessments.

After selection of the initial sample of schools was completed, information was obtained to update the sample for new eligible schools. Public school districts and Catholic dioceses of initially selected schools were asked to give information about new schools and schools with changes in grade structure since the 1988 date to which the CCD and QED lists related, for their district/diocese. Schools so identified were given an appropriate chance of inclusion in each of the samples for which they were eligible. The overall probability of inclusion for a given age class for each such school was determined by the estimated number of eligible students enrolled in the school, and the within-PSU sampling rate used to select the initial sample of schools. The conditional probability of selection, used to draw the actual samples of new schools, was based upon the selection probabilities of those schools from the district/diocese initially selected from the frame. This procedure, described in detail in Rust et al. (1992), when combined with an appropriate weighting scheme, ensured unbiased representation of such new schools throughout the United States. The process identified a total of 172 new schools, 140 eligible for age class 9, 119 eligible for age class 13, and 36 eligible for age class 17. Four new schools were added to the main samples in this way—three at age class 9 and one at age class 17. All four schools were found to have eligible students enrolled, and were invited to participate.

In a few PSUs where school refusals were relatively heavy for a particular sample, substitute school selections were made, replacing the refusals (to the extent feasible) with schools from within the same PSU and similar in size, affiliation (public, Catholic, or other private), grade span, and minority composition. The goal of this procedure was to maintain the student sample sizes needed, while keeping variance and nonresponse bias at acceptable levels. Table 3-5 shows the number of in-scope schools selected, cooperating, and substituted, in each of the school samples. The participation rates given are based on the initially selected sample of schools. These response rates are comparable with those of assessments conducted during the 1980s. Note that since the response rates quoted do not include the substitute selections, the potential for nonresponse bias is likely to be a little less than these rates would indicate. This is because the substitute selections were chosen based on their similarity to the initially refusing selections.

Table 3-5
School Sample Sizes, Refusals, and Substitutes for the Main Samples

	Age Class 9	Age Class 13	Age Class 17	Total	Public*	Private**
Selected, in scope	589	520	417	1,526	945	593
Refusals	69	69	78	216	124	92
Participation rate of originally selected schools	88%	87%	81%	86%	87%	84%
1988 participation rate	89%	87%	83%	86%	—	—
Participating, no eligibles enrolled	12	49	38	99	44	55
Substitutes participating	7	4	3	14	13	1
Final assessed sample	527	406	304	1,237	790	447

* Public, Bureau of Indian Affairs, and Department of Defense schools
** Catholic and other private schools

The considerable numbers of schools selected with no eligible students enrolled resulted primarily from the fact that, for example, for age 13/grade 8, some schools with grades 6, 7, or 9, but no grade 8, were sampled. Such schools had a reasonable chance of containing some age 13 students. Often they did have a number of eligible students, but sometimes they had none. Because of the grade structure of schools, this occurred most often for age 13/grade 8.

A school characteristics and policies questionnaire was mailed to every sampled school by Westat before the assessment. The Westat supervisor then collected the questionnaires and returned them to ETS. The school characteristics and policies questionnaire is described in Chapter 4.

A school principal's questionnaire, distributed to each sampled school by Westat before the assessment, was used to refine the estimate of the age/grade-eligible students and to determine in part the size and type of community (STOC) codes (see Rust et al., 1992).

Three different session types were administered at each age class. One was a spiraled, print-administered session, in which some students were assessed in reading while others were assessed in science. The second was a spiraled, print-administered mathematics session. The third involved a tape-administered mathematics booklet. These three session types were assigned among the selected schools found to be in-scope, and, at the time of assigning sessions, to schools that were likely or possible participants in NAEP.

First, the minimum proportion of sampled students within a school who could be assigned to a single session, without that session's being unduly small, was established. Thus, for schools with few eligible students, all students were to be assigned to a single session, to be of only one of the three session types above. In large schools at age classes 13 and 17, where it was anticipated that 200 students would be selected, this proportion was set at one-ninth. For

large schools at age class 9, where it was anticipated that 150 students would be selected, this proportion was set at one-sixth. Intermediate proportions were set for schools in intermediate size (see Rust et al., 1992). Session types were then assigned to schools with three aims in mind. The first was to distribute students to the different session types, across the whole sample (and so far as possible, proportionately within each PSU), for each age class, so that the target numbers of assessed students would be achieved. The second was to maximize the number of different session types that were administered within a given selected school, without violating the minimum proportions within a single session (discussed above). The third was to give each student selected for the main sample for an age class an equal chance of being selected for a given session type. Thus, for example, large schools at age classes 13 and 17 had one-ninth of their selected students assigned to the mathematics tape session, and either two- or three-ninths assigned to the mathematics spiral session, with the remainder assigned to the reading/science spiral session. Overall, and as closely as possible within each PSU, 29.6 percent were assigned to mathematics spiral sessions and 59.3 percent to reading/science sessions. Large schools at age class 9 had a 60 percent chance of having one-sixth of their students assigned to the mathematics tape session, and a 40 percent chance of not conducting a mathematics tape session at all. Either 16.7 percent, 33.3 percent, or 40 percent of the students in such schools were assigned to the mathematics spiral sessions, in such a way that overall, and as far as possible within each PSU, 26.7 percent of students were assigned to mathematics spiral sessions and 63.3 percent to reading/science.

On occasion, a school failed to participate after having been assigned its allocation of session types. Often also such a school was not replaced by a substitute selection. Thus two types of school level nonresponse, designated as school and session, were needed to address the consequences of refusal prior to, and after, session assignments were made. Adjustments to the sample weights were made for each of these nonresponse components, as discussed in Chapter 10.

The procedure was intended to ensure that each session type was conducted within each PSU, and to the extent feasible, within each season within each PSU. The relatively small proportion of age class 9 schools assigned to conduct the mathematics tape session, and the fact that most school level nonresponse occurred after the assignment of sessions, meant that, especially at the season-specific level, this was not always possible to achieve, especially for age class 9. The use of this procedure, however, helped to ensure that the different session types were spread among PSUs and seasons to the maximum extent feasible in practice.

3.3 SELECTION OF SCHOOLS FOR BRIDGE SAMPLES; THE ASSIGNMENT OF SESSIONS TO SCHOOLS

Bridge sample schools were selected for each age class from all 97 PSUs. This was a change from 1988 and earlier years, when bridge samples were selected from a subset of PSUs. The administrative procedures in the field for 1990, involving both the national and state samples, and the relatively large sample sizes for the 1990 bridges were such that this undesirable step of further clustering the bridge sample schools into a small set of PSUs was not necessary.

The sample of schools was drawn for the bridge samples in a manner very similar to that used for the main samples. The differences were, first, that no subgroups of schools (high-minority enrollment or private schools) were identified for oversampling (though small schools were still undersampled); second, as explained above, the special supplement to the private school frame was not utilized; third, the minimum sample sizes of schools to be selected per PSU were set at three for each age class, regardless of the rate of sampling of schools for NELS; fourth, the within-PSU probability of selection for any school in a given age class was capped at 0.5, to ensure that adequate schools remained to be selected for the main sample. Finally, the measure of size used for each school was the estimated number of age eligible students in the school, since for each age class the large majority of students selected were assigned to sessions for only students of the appropriate age were eligible. The maximum size of the school in which all age-eligible students would be asked to participate was set at 120 for each age class. In most schools having the modal grade, some additional students were selected who were in the modal grade but not age eligible, so that the maximum sample size of students within a school was about 160 grade and age eligible students.

Substitute selections were made for nonparticipating bridge sample schools for the age 13/grade 8 bridge sample in certain PSUs in a manner similar to the approach used for the main sample schools. No substitutes were chosen for the other bridge samples, because the distribution and timing of school refusals made substitution both difficult and not very necessary. As in the case of the main samples, samples of new schools were selected. For each age class, one school was selected to be added to the sample in this manner.

Table 3-6 shows the school sample sizes and participation rates for the bridge samples for each age class. School participation rates are similar to those seen in equivalent samples in 1988—the fall and winter bridges and the age 17/grade 11 spring bridge sample.

Table 3-6
School Sample Sizes, Refusals, and Substitutes for the Bridge Samples

	Age 9/Grade 4 (Winter)	Age 13/Grade 8 (Fall)	Age 17/Grade 11 (Spring)	Total
Selected, in scope	335	368	373	1,076
Refusals	40	36	72	148
Participation rate of originally selected schools	88%	90%	81%	86%
1988 participation rate	87%	93%	78%	87%
Participating, no eligibles enrolled	12	50	9	71
Substitutes participating	5	0	0	5
Final assessed sample	288	282	292	862

For all three age classes, sessions were assigned to bridge sample schools in the following manner. First, the number of sessions per school was established. This was the maximum number of sessions (up to four) that could be administered without creating unduly small session

sizes with few eligibles. Thus, in most bridge sample schools, four sessions were conducted. However, schools with fewer than 20 eligibles, for example, were asked to conduct only a single session.

The number of session types conducted in the assessment varied by age class. Table 3-7 in the following section shows, among other things, the various bridge sample session types conducted for each age class, and the year of the corresponding assessment to which these session types provided a bridge.

The assignment of sessions to schools maximized the number of session types conducted within each PSU. Thus, to the extent feasible, session assignment was delayed until after it was determined that a selected school would participate. Just as for the main sample, on a few occasions a session could not be conducted in a school that, at the time of session assignment, was expected to participate but subsequently did not. As a result, two types of school nonresponse adjustment factors, denoted school and session, were required for the bridge samples (see Chapter 10).

This procedure was intended to assure that each session type was assigned to the maximum number of PSUs feasible, given practical constraints in the field. Since it was not feasible to administer each of the various five or six different session types in a single school, not all session types were administered in all 97 PSUs, but each was administered in all but a handful of PSUs.

3.4 SAMPLING STUDENTS

In the fourth stage of sampling, a consolidated list was prepared for each school of all grade-eligible and age-eligible students for the age class for which the school was selected. A systematic selection of eligible students was made from this list (unless all students were to be assessed) to provide the target sample size. For schools assigned to more than a single session type (the vast majority), students were assigned by Westat district supervisors to one of the various session types using specified procedures. In the bridge samples, students assigned to paced-tape sessions who were not age-eligible were dropped from the assessment.

For each age class, separately for the bridge and main samples, maxima were established as to the number of students who would be selected for a given school. In those schools that, according to information on the frame, had fewer eligible students than the established maxima, each eligible students enrolled at the school was selected in the sample for one of the sessions assigned to the school. In other schools, a sample of students was drawn, and then students were assigned to sessions as appropriate. For the main samples, the maximum sample sizes were established in terms of the number of grade- plus age-eligible students—150 at age class 9 and 200 at age classes 13 and 17. For the bridge samples, the maximum at each age class was 160 grade- plus age-eligible students. Note that the number of students actually selected for assessment in a bridge sample school generally fell somewhat below 160, because students who were selected for one of the bridge tape sessions and were in the modal grade but not age-eligible were subsequently dropped from the sample.

The sample for students to be selected in each school was derived in the following manner, both for main and for bridge samples. On the basis of data obtained from the principal questionnaire (or the sample frame when the principal questionnaire data were not obtained in time) an estimate of the number of eligible students was established for each school. For the main samples, the estimated number of grade- plus age-eligible students was used; for the bridge samples, the number of age-eligible students was used. A Session Assignment Form was generated for each school, showing the line numbers (described below) of the students to be selected, indicating the type of session to be taken by each such student. These line numbers were generated using a sampling interval designed to give the appropriate sample size for each school. Thus the overall sampling interval was 1.0 for schools in which all eligible students were to be assessed. The appropriate sampling interval was specified for schools with larger numbers of eligible students, such as to give the appropriate maximum sample size (described above for each age class) in the case that the school had an enrollment of eligible students exactly equal to that predicted.

If the Westat supervisor found that, when applied to the numbered list of eligible students assembled in the field for each school, the line numbers generated gave rise to a sample in excess of 120 percent of the appropriate maximum sample size limit specified above, he or she called Westat's central office. By use of a personal computer, new line numbers based on the actual number of eligible students were generated and relayed to the supervisor. A similar revision to the line numbers was made in the case of a school with a sampling interval in excess of 1.0, and eligible enrollment less than 80 percent of that initially estimated. In this latter case the sample size was increased to the appropriate level. This procedure gave a suitable compromise between control over the sampling rate within each school and operational autonomy and flexibility for Westat field supervisors. Note that in all cases, sampling intervals were generated in Westat's central office, and stored for use in sample weighting. Supervisors were not required to derive or record within-school sampling rates.

Table 3-7 shows the number of students per school who were assessed for each session type. Note that, for the various spiral samples, the number of students assessed per item per school is quite low, even though typically dozens of students were assessed in total in a particular school. Thus the extent of clustering of the sample is in general quite modest, because most sampled schools conducted a few different types of sessions with a moderate number of students in each, and more importantly because the use of BIB spiraling in the print-administered sessions greatly alleviated the effects of clustering the samples of students within schools.

3.5 EXCLUDED STUDENTS

Some students selected for the sample were deemed unassessable by school authorities because they had limited English language proficiency, were judged as being mildly mentally retarded (educable), or were functionally disabled. In these cases, school staff completed an excluded student questionnaire, listing the reason for exclusion and providing some background information.

Nine distinct samples of excluded students were identified. For each age class, there was one sample for the bridge assessment, one for the main sample winter assessment, and one for

Table 3-7
Number of Students per School for Each Session Type

Sample	Sample Code	Session Type	Number of Schools	Mean Number of Students per Session Type per School	Mean Number of Students per Item per School
Age Class 9 Bridges	RW-Br84	Spiral booklets 51-56	224	23.6	3.9 - 7.8*
	RMS-Br86	Tape booklet 91	122	16.3	16.3
	RMS-Br86	Tape booklet 92	120	17.1	17.1
	RMS-Br86	Tape booklet 93	126	17.4	17.4
	MS-BrLT	Tape booklet 94	123	16.7	16.7
	MS-BrLT	Tape booklet 95	120	17.4	17.4
Age Class 9 Main	Rdg,Sci-MainP	Spiral Reading/Science	498	41.2	8.8
	Math-MainP	Spiral Mathematics	372	23.6	10.1
	Math-MainT	Tape Mathematics	165	19.3	19.3
Age Class 13 Bridges	RW-Br84	Spiral booklets 51-56	227	27.5	4.1 - 8.2*
	RMS-Br86	Tape booklet 91	114	19.6	19,6
	RMS-Br86	Tape booklet 92	117	19.6	19.6
	RMS-Br86	Tape booklet 93	109	19.6	19.6
	MS-BrLT	Tape booklet 94	118	19.1	19.1
	MS-BrLT	Tape booklet 95	121	18.2	18.2
Age Class 13 Main	Rdg,Sci-MainP	Spiral Reading/Science	353	49.4	10.6
	Math-MainP	Spiral Mathematics	288	30.0	12.9
	Math-MainT	Tape Mathematics	185	17.2	17.2
Age Class 17 Bridges	RW-Br84	Spiral booklets 51-56	227	24.7	4.1 - 8.2*
	RMS-Br86	Spiral booklets 61-66	261	31.9	5.3 - 15.9*
	MS-Br86	Tape booklet 84	112	19.7	19.7
	MS-Br86	Tape booklet 85	112	19.7	19.7
	MS-BrLT	Tape booklet 94	108	20.3	20.3
	MS-BrLT	Tape booklet 95	114	19.4	19.4
Age Class 17 Main	Rdg,Sci-MainP	Spiral Reading/Science	278	60.4	12.9
	Math-MainP	Spiral Mathematics	255	33.0	14.1
	Math-MainT	Tape Mathematics	196	16.0	16.0

* This number varied because some item blocks appeared more than once in the set of booklets used for this sample (see Chapter 4).

the main sample spring assessment. For many purposes of analysis, the two main sample components could be combined, as they represent the same population, and used identical exclusion criteria. The exclusion criteria for the main samples differed somewhat for those used for the bridge samples and in previous years. The exclusion criteria for the main samples were identical to those used for the 1990 Trial State Assessment, and were intended to be somewhat more rigorously defined than those used in the bridge samples. (For more details of the exclusion criteria and their implementation, see Chapter 5.) In addition, for age class 17, the excluded students from the bridge assessments (with an October-September age definition and modal grade of 11) were drawn from a population different from that of the excluded students in the main assessment (with a calendar-year age definition and modal grade of 12).

For all samples, students were selected for specific sessions, and the school was then asked to identify those to be excluded. Thus only age-eligible students were considered for exclusion from the bridge tape sessions, whereas both age- and grade-eligible students were considered for exclusion from the main samples and the spiral bridge samples. The samples of excluded students for the bridge samples were weighted in such a way as to account for this procedure appropriately (see Chapter 10).

Table 3-8 shows the rates of exclusion for each age class for the bridge and main samples. For the main samples, for which private school students were oversampled by a factor of three, and constituted about 18 percent of the student sample, exclusion rates are shown for both public and private schools. Overall rates for 1988 (when no oversampling of private schools took place) are shown for comparison. The table shows that rates of exclusion are very similar for the main and bridge samples. Somewhat greater rates of exclusion have occurred in the main than in the bridge samples, at least within public schools, but this is somewhat masked by the presence of higher proportions of private school students in the main samples. Exclusion rates also appear to have increased slightly over those observed in 1988. The most marked effects, however, are the much higher rates of exclusion in public schools than in private, and the higher rates of exclusion at lower grades. The former phenomenon is no doubt a function of the greater prevalence of special education and language minority programs in public schools. The higher exclusion rates at lower ages, which occurred also in 1988, results from the greater proportion of students at these grades who are excluded for reasons of limited English proficiency. In certain areas of the United States, fourth-grade public-school students whose native language is Spanish are taught predominantly in Spanish, and in these schools a very high proportion of sampled students are excluded.

Table 3-8
Student Exclusion Rates by Age Class and School Type, Unweighted

Samples	1990 Excluded (%)			1988 Excluded (%)
	Public	Private	Total	Total
Age 9/Grade 4 Bridge	—	—	6.1	6.3*
Age 9/Grade 4 Main	7.7	0.8	6.5	
Age 13/Grade 8 Bridge	—	—	5.5	5.3*
Age 13/Grade 8 Main	7.2	0.9	6.1	
Age 17/Grade 11 Bridge	—	—	4.4	3.0
Age 17/Grade 12 Main	4.9	0.9	4.2	3.7*

* Somewhat different exclusion criteria were used for the 1990 main samples than for the 1990 bridge samples and the 1988 samples. Note also that the total rates for 1990 main samples are based on a relatively greater contribution from private school students. Private school students constitute about 18 percent of the sample for the 1990 main samples, and about 7 percent of the sample for the 1990 bridge samples and the 1988 samples.

3.6 STUDENT PARTICIPATION RATES

Table 3-9 summarizes the rates of participation of invited students. The set of invited students consists of the selected students, after removing the excluded students and, in the case of bridge samples, removing those students selected for tape sessions who were not age eligible. For a given session, a makeup session was called for when, for various reasons, more than a tolerable number of invited students failed to attend the originally scheduled session to which they were invited. The participation rates given in the table express the number finally assessed as a percentage of those initially invited in the participating schools. Participation rates are shown for the main and bridge samples (combined across winter and spring seasons for the main samples), and for public and private schools separately in the case of the main samples. Overall participation rates are also shown for comparable samples from the 1988 NAEP assessment. The table shows that student participation rates in 1990 are very similar to those experienced in 1988 for age classes 9 and 13 with an improvement of about two percentage points evident at age class 17, both for the bridge sample, consisting predominantly of eleventh graders, and the main samples, largely made up of twelfth graders. At all age classes, the participation rate of private school students exceeds that of public school students, with the difference, both relative and absolute, increasing with age class. This is in contrast with the levels of school participation, which are quite similar for public and private schools at each age class.

Table 3-9
Student Participation Rates by Age Class and School Type, Unweighted

Samples	1990 Public		1990 Private		1990 Total		1988 Participation Rate
	Number Invited	Participation Rate (%)	Number Invited	Participation Rate (%)	Number Invited	Participation Rate (%)	
Age 9/Grade 4 Bridge	—	—	—	—	17,626	92.4	92.2
Age 9/Grade 4 Main	28,454	92.5	6,520	94.6	34,974	92.9	92.8*
Age 13/Grade 8 Bridge	—	—	—	—	19,187	90.4	90.0
Age 13/Grade 8 Main	27,121	88.0	5,720	94.3	32,841	89.1	87.8*
Age 17/Grade 11 Bridge	—	—	—	—	27,778	81.2	79.2
Age 17/Grade 12 Main	28,383	80.0	6,492	87.3	34,875	81.3	78.5*

* The total rates for the 1990 main samples are based on a relatively greater contribution of private school students than either the 1988 samples or the 1990 bridge samples. Private school students constitute about 18 percent of the invited students for the 1990 main samples, and about 7 percent of the invited students for the 1990 bridge samples and the 1988 samples.

3.7 OVERALL STUDENT PARTICIPATION RATES

The combined impact of school nonparticipation and student absenteeism from sessions within participating schools is summarized in Table 3-10. The table shows the percentages of students assessed, from among those who would have been assessed if all initially selected schools had participated, and if all invited students had attended either an initial or make-up session. The results show that, consistent with earlier rounds of NAEP, the overall level of participation decreases substantially with the increase in age and grade of the students.

The procedures for substituting for nonparticipating schools or imputing for them and the procedures for imputing for absent students were designed (so far as feasible) to reduce the biases resulting from school and student nonparticipation. These procedures are discussed in Chapter 10.

3.8 SAMPLING TEACHERS

The teacher questionnaire was administered to the mathematics teachers of fourth-grade and eighth-grade students sampled for the main assessment who were assessed in either the mathematics spiral or mathematics tape sessions and to the science teachers of eighth-grade students sampled for the main assessment who were assessed in a reading/science spiral using a science booklet. The purpose of drawing these samples was not to estimate the attributes of the teacher population, but to estimate the number (proportion) of students whose teachers had various attributes and to correlate student characteristics and performance with the characteristics of their teachers.

Table 3-10
Overall Participation Rates (School and Student Combined) by Age Class

1990 Samples	Age Class 9	Age Class 13	Age Class 17	Overall
Main Samples				
School participation	88.3%	86.7%	81.3%	85.8%
Student participation	92.9%	89.1%	81.3%	87.7%
Overall student participation	82.0%	77.2%	66.1%	74.8%
Number of participating students	32,490	29,250	28,341	90,081
Bridge Samples				
School participation	88.1%	90.5%	80.7%	86.3%
Student participation	92.4%	90.4%	81.2%	87.0%
Overall student participation	81.4%	81.8%	65.5%	74.2%
Number of participating students	16,295	17,337	22,765	56,397
Overall				
School participation	88.2%	88.3%	81.0%	86.0%
Student participation	92.7%	89.6%	81.3%	87.4%
Overall student participation	81.8%	78.9%	65.8%	74.6%
Number of participating students	48,785	46,587	51,106	146,478

All such teachers were included in the sample, and were asked to complete a questionnaire concerning themselves and their teaching practices, with specific references to each individual class period containing a student included in the main assessment. Only teachers of students taking fourth-grade mathematics, eighth-grade mathematics or eighth-grade science were included. Thus, for example, the teacher of an eighth-grade student assessed in mathematics who was taking ninth- or seventh-grade mathematics was not included in the sample.

Chapter 4

ASSESSMENT INSTRUMENTS[1]

Debra L. Kline

Educational Testing Service

In the 1990 assessment, five types of instruments were used to collect data about students, teachers, and schools. Each assessed student received an **assessment booklet** containing both cognitive and background questions. An **excluded student questionnaire** was completed by school officials for each sampled student who was deemed unable to take part in the assessment. **Teacher questionnaires** were given to the fourth-grade mathematics, eighth-grade mathematics, or eighth-grade science teachers of fourth- or eighth-grade students who took part in the main assessments of mathematics or science. A **school characteristics and policies questionnaire** was distributed to each participating school. A **principal questionnaire** was completed prior to the assessment for each school selected for the sample.

In addition, as part of a special study of school-based writing, NAEP collected writing papers from random samples of students at age 9/grade 4 and age 13/grade 8.

This chapter begins with a discussion of the characteristics of the student booklets used for the 1990 main and bridge assessments and how the booklets were assembled. A set of tables presents in detail the contents of each booklet and item block. Sections 4.3 through 4.6 provide an overview of the contents of the excluded student, teacher, school characteristics and policies, and principal questionnaires. Section 4.7 describes the special study that was part of the writing trend assessment.

4.1 Student Booklets—Main Assessments

As described in Chapter 2, segments or "blocks" of multiple-choice and open-ended cognitive items were created from the pool of items for each of the assessed subject areas. Ten blocks of mathematics items (labeled MC - ML), seven blocks of reading items (RC - RI), and seven blocks of science items (SC - SI) were created for each age/grade level in the 1990 main assessments.

One mathematics block (MF) at age 9/grade 4 and age 13/grade 8 contained items that required the use of a ruler or protractor; two blocks (MH and MI) at all three levels had items

[1] The author acknowledges Lynn Jenkins and the authors of Chapter 2 for their contributions to this chapter.

that required the use of a calculator. Science block SH at all three levels contained figural response questions in which students were required to draw arrows or lines on figures and graphs in response to particular tasks.

Three of the 10 mathematics blocks (MJ, MK, and ML) were composed of items designed to assess students' skills in estimation or higher-order thinking. These three blocks were assembled with a fourth (non-estimation) mathematics block (MD) into one booklet for each age/grade level. This special study booklet was accompanied by a paced audiotape when it was administered to the students.

The seven non-estimation mathematics blocks and the seven reading and science blocks were assembled into subject-area booklets according to the focused-BIB spiral method (see Chapter 1), resulting in seven different booklets in each subject area for each age/grade.

In each booklet in the main assessments, the cognitive blocks were preceded by a student demographics questionnaire (block CA) and one of three student content area questionnaires (block RB for reading booklets, MB for mathematics booklets, and SB for science booklets). Although many of the items in these questionnaires were common to the three age/grade levels, some were specifically targeted to a particular level. The students demographics questionnaire included questions about race/ethnicity, language spoken in the home, reading materials in the home, mother's and father's level of education, homework, academic expectations, which parents live at home, and which parents work. The student content area questionnaires were specific to each subject area and included questions about instructional activities, courses taken, use of specialized resources, and views about the utility and value of the subject matter.

Tables 4-1, 4-2, and 4-3 summarize the contents of each main assessment booklet and show how many of each booklet were administered. Tables 4-4, 4-5, and 4-6 give details of the item blocks used in the main assessments, including the number of cognitive and open-ended items in each block and the booklets in which each block appeared.

Answer Documents

In NAEP assessments prior to 1986, students circled answers (for multiple-choice questions) or wrote responses (for open-ended questions) directly in their assessment booklets. These answers were later manually transcribed by NAEP staff. In 1986, because of an increase in the volume of response data and a shorter time period for processing, a new type of booklet was introduced in which students answered multiple-choice questions by filling in ovals next to the responses in the booklets, which were then fed through an optical scanner. In 1988, to further improve the efficiency of NAEP data processing, students in the two older age classes were given a separate, scannable answer sheet on which to record their responses both to multiple-choice and to open-ended questions.

In the 1990 main assessments, two types of answer documents were used. Scannable booklets were used for the mathematics assessment; separate, scannable answer sheets were used for the reading assessment. For the science assessment, only the figural response items in block SH were answered directly in the booklet; students recorded answers to all other questions on a separate answer sheet.

Table 4-1

Main Sample Booklet Contents and Number of Booklets Administered
Age 9/Grade 4

MAIN SAMPLE, AGE 9/GRADE 4						
Subject Area	**Booklet Number**	**Answer Document**[1]	**Common Background Block**	**Subject Area Background Block**	**Cognitive Blocks**	**Number of Booklets Administered**
Reading	1	A	CA	RB	RC RD RF	1,199
	2	A	CA	RB	RD RE RG	1,214
	3	A	CA	RB	RE RF RH	1,207
	4	A	CA	RB	RF RG RI	1,202
	5	A	CA	RB	RG RH RC	1,217
	6	A	CA	RB	RH RI RD	1,223
	7	A	CA	RB	RI RC RE	1,218
Answer Book Bridge	8	B	CB	RA	RC RD RL	1,205
	9	B	CB	RA	RK RE RC	1,206
	10	B	CB	RA	RE RJ RD	1,204
Mathematics	11	B	CA	MB	MC MD MF[2]	1,255
	12	B	CA	MB	MD ME MG	1,250
	13	B	CA	MB	ME MF[2] MH[3]	1,261
	14	B	CA	MB	MF[2] MG MI[3]	1,242
	15	B	CA	MB	MG MH[3] MC	1,263
	16	B	CA	MB	MH[3] MI[3] MD	1,260
	17	B	CA	MB	MI[3] MC ME	1,259
Science	18	A	CA	SB	SC SD SF	1,221
	19	A	CA	SB	SD SE SG	1,207
	20	A/B	CA	SB	SE SF SH[4]	1,208
	21	A	CA	SB	SF SG SI	1,204
	22	A/B	CA	SB	SG SH[4] SC	1,197
	23	A/B	CA	SB	SH[4] SI SD	1,202
	24	A	CA	SB	SI SC SE	1,179
Mathematics Estimation and Higher Order Thinking Skills	28	B	CA	MB	MD MJ MK ML	3,187

[1] A = separate, scannable answer sheet; B = scannable booklet. For booklets 20, 22, and 23, block SH required answers in the booklet; the remainder of those booklets required answers on a separate answer sheet.

[2] Ruler needed for this block

[3] Calculator needed for this block

[4] Figural responses required for this block

Table 4-2

Main Sample Booklet Contents and Number of Booklets Administered
Age 13/Grade 8

MAIN SAMPLE, AGE 13/GRADE 8						
Subject Area	**Booklet Number**	**Answer Document**[1]	**Common Background Block**	**Subject Area Background Block**	**Cognitive Blocks**	**Number of Booklets Administered**
Reading	1	A	CA	RB	RC RD RF	1,263
	2	A	CA	RB	RD RE RG	1,257
	3	A	CA	RB	RE RF RH	1,248
	4	A	CA	RB	RF RG RI	1,250
	5	A	CA	RB	RG RH RC	1,229
	6	A	CA	RB	RH RI RD	1,238
	7	A	CA	RB	RI RC RE	1,240
Mathematics	8	B	CA	MB	MC MD MF[2]	1,234
	9	B	CA	MB	MD ME MG	1,244
	10	B	CA	MB	ME MF[2] MH[3]	1,230
	11	B	CA	MB	MF[2] MG MI[3]	1,240
	12	B	CA	MB	MG MH[3] MC	1,225
	13	B	CA	MB	MH[3] MI[3] MD	1,236
	14	B	CA	MB	MI[3] MC ME	1,225
Science	15	A	CA	SB	SC SD SF	1,233
	16	A	CA	SB	SD SE SG	1,243
	17	A/B	CA	SB	SE SF SH[4]	1,246
	18	A	CA	SB	SF SG SI	1,234
	19	A/B	CA	SB	SG SH[4] SC	1,245
	20	A/B	CA	SB	SH[4] SI SD	1,248
	21	A	CA	SB	SI SC SE	1,260
Mathematics Estimation and Higher Order Thinking Skills	25	B	CA	MB	MD MJ MK ML	3,182

[1] A = separate, scannable answer sheet; B = scannable booklet. For booklets 17, 19, and 20, block SH required answers in the booklet; the remainder of those booklets required answers on a separate answer sheet.

[2] Ruler needed for this block

[3] Calculator needed for this block

[4] Figural responses required for this block

Table 4-3

Main Sample Booklet Contents and Number of Booklets Administered
Age 17/Grade 12

MAIN SAMPLE, AGE 17/GRADE 12									
Subject Area	Booklet Number	Answer Document[1]	Common Background Block	Subject Area Background Block	Cognitive Blocks				Number of Booklets Administered
Reading	1	A	CA	RB	RC	RD	RF		1,180
	2	A	CA	RB	RD	RE	RG		1,185
	3	A	CA	RB	RE	RF	RH		1,196
	4	A	CA	RB	RF	RG	RI		1,196
	5	A	CA	RB	RG	RH	RC		1,191
	6	A	CA	RB	RH	RI	RD		1,196
	7	A	CA	RB	RI	RC	RE		1,207
Mathematics	8	B	CA	MB	MC	MD	MF		1,201
	9	B	CA	MB	MD	ME	MG		1,201
	10	B	CA	MB	ME	MF	MH[2]		1,193
	11	B	CA	MB	MF	MG	MI[2]		1,215
	12	B	CA	MB	MG	MH[2]	MC		1,187
	13	B	CA	MB	MH[2]	MI[2]	MD		1,211
	14	B	CA	MB	MI[2]	MC	ME		1,198
Science	15	A	CA	SB	SC	SD	SF		1,200
	16	A	CA	SB	SD	SE	SG		1,215
	17	A/B	CA	SB	SE	SF	SH[3]		1,223
	18	A	CA	SB	SF	SG	SI		1,201
	19	A/B	CA	SB	SG	SH[3]	SC		1,215
	20	A/B	CA	SB	SH[3]	SI	SD		1,200
	21	A	CA	SB	SI	SC	SE		1,191
Mathematics Estimation and Higher Order Thinking Skills	25	B	CA	MB	MD	MJ	MK	ML	3,139

[1] A = separate, scannable answer sheet; B = scannable booklet. For booklets 17, 19, and 20, block SH required answers in the booklet; the remainder of those booklets required answers on a separate answer sheet.
[2] Calculator needed for this block
[3] Figural responses required for this block

Table 4-4

Main Sample Block Information, Age 9/Grade 4

Block	Type	Total Number of Items	Number of Cognitive Items	Number of Open-ended Items	Booklets Containing Block
CA	Common Background	18	0	0	1 - 7, 11 - 24, 28
CB	Common Background	21	0	0	8 - 10
RA	Reading Background	10	0	0	8 - 10
RB	Reading Background	10	0	1	1 - 7
RC	Reading Cognitive	7	7	0	1, 5, 7, 8, 9
RD	Reading Cognitive	7	7	0	1, 2, 6, 8, 10
RE	Reading Cognitive	12	12	1	2, 3, 7, 9, 10
RF	Reading Cognitive	9	9	1	1, 3, 4
RG	Reading Cognitive	12	12	0	2, 4, 5
RH	Reading Cognitive	11	11	0	3, 5, 6
RI	Reading Cognitive	9	9	1	4, 6, 7
RJ	Reading Cognitive	11	11	0	10
RK	Reading Cognitive	14	13	0	9
RL	Reading Cognitive	15	15	6	8
MB	Mathematics Background	14	0	0	11 - 17, 28
MC	Mathematics Cognitive	19	19	6	11, 15, 17
MD	Mathematics Cognitive	14	14	0	11, 12, 16, 28
ME	Mathematics Cognitive	11	11	11	12, 13, 17
MF	Mathematics Cognitive (Ruler)	17	17	4	11, 13, 14
MG	Mathematics Cognitive	18	18	0	12, 14, 15
MH	Mathematics Cognitive (Calculator)	15	15	1	13, 15, 16
MI	Mathematics Cognitive (Calculator)	15	15	6	14, 16, 17
MJ	Mathematics Cognitive (Estimation)	20	20	0	28
MK	Mathematics Cognitive (Higher-order Skills)	7	7	7	28
ML	Mathematics Cognitive (Higher-order Skills)	7	7	6	28
SB	Science Background	15	0	0	18 - 24
SC	Science Cognitive	17	17	0	18, 22, 24
SD	Science Cognitive	21	21	0	18, 19, 23
SE	Science Cognitive	19	19	0	19, 20, 24
SF	Science Cognitive	13	13	4	18, 20, 21
SG	Science Cognitive	20	20	0	19, 21, 22
SH	Science Cognitive (Figural Response)	10	10	10	20, 22, 23
SI	Science Cognitive	12	12	3	21, 23, 24

Table 4-5

Main Sample Block Information, Age 13/Grade 8

Block	Type	Total Number of Items	Number of Cognitive Items	Number of Open-ended Items	Booklets Containing Block
CA	Common Background	22	0	0	1 - 21, 25
RB	Reading Background	18	0	1	1 - 7
RC	Reading Cognitive	9	9	0	1, 5, 7
RD	Reading Cognitive	10	10	0	1, 2, 6
RE	Reading Cognitive	19	19	0	2, 3, 7
RF	Reading Cognitive	14	14	0	1, 3, 4
RG	Reading Cognitive	15	15	2	2, 4, 5
RH	Reading Cognitive	15	15	0	3, 5, 6
RI	Reading Cognitive	15	15	0	4, 6, 7
MB	Mathematics Background	22	0	0	8 - 14, 25
MC	Mathematics Cognitive	23	23	4	8, 12, 14
MD	Mathematics Cognitive	21	21	0	8, 9, 13, 25
ME	Mathematics Cognitive	16	16	16	9, 10, 14
MF	Mathematics Cognitive (Protractor/Ruler)	21	21	5	8, 10, 11
MG	Mathematics Cognitive	18	18	1	9, 11, 12
MH	Mathematics Cognitive (Calculator)	18	18	2	10, 12, 13
MI	Mathematics Cognitive (Calculator)	20	20	7	11, 13, 14
MJ	Mathematics Cognitive (Estimation)	22	22	0	25
MK	Mathematics Cognitive (Estimation)	24	24	0	25
ML	Mathematics Cognitive (Higher-order Skills)	8	8	7	25
SB	Science Background	25	0	0	15 - 21
SC	Science Cognitive	22	22	0	15, 19, 21
SD	Science Cognitive	26	26	0	15, 16, 20
SE	Science Cognitive	26	26	0	16, 17, 21
SF	Science Cognitive	18	18	4	15, 17, 18
SG	Science Cognitive	23	23	3	16, 18, 19
SH	Science Cognitive (Figural Response)	14	14	14	17, 19, 20
SI	Science Cognitive	17	17	3	18, 20, 21

Table 4-6

Main Sample Block Information, Age 17/Grade 12

Block	Type	Total Number of Items	Number of Cognitive Items	Number of Open-ended Items	Booklets Containing Block
CA	Common Background	30	0	0	1 - 21, 25
RB	Reading Background	25	0	1	1 - 7
RC	Reading Cognitive	18	18	0	1, 5, 7
RD	Reading Cognitive	15	15	1	1, 2, 6
RE	Reading Cognitive	19	19	0	2, 3, 7
RF	Reading Cognitive	14	14	0	1, 3, 4
RG	Reading Cognitive	14	14	1	2, 4, 5
RH	Reading Cognitive	17	17	0	3, 5, 6
RI	Reading Cognitive	15	15	1	4, 6, 7
MB	Mathematics Background	34	0	0	8 - 14, 25
MC	Mathematics Cognitive	23	23	5	8, 12, 14
MD	Mathematics Cognitive	22	22	0	8, 9, 13, 25
ME	Mathematics Cognitive	17	17	17	9, 10, 14
MF	Mathematics Cognitive	20	20	3	8, 10, 11
MG	Mathematics Cognitive	21	21	3	9, 11, 12
MH	Mathematics Cognitive (Calculator)	21	21	4	10, 12, 13
MI	Mathematics Cognitive (Calculator)	20	20	3	11, 13, 14
MJ	Mathematics Cognitive (Estimation)	22	22	0	25
MK	Mathematics Cognitive (Estimation)	24	24	0	25
ML	Mathematics Cognitive (Higher-order Skills)	13	13	13	25
SB	Science Background	32	0	0	15 - 21
SC	Science Cognitive	25	25	0	15, 19, 21
SD	Science Cognitive	29	29	0	15, 16, 20
SE	Science Cognitive	24	24	0	16, 17, 21
SF	Science Cognitive	17	17	1	15, 17, 18
SG	Science Cognitive	26	26	3	16, 18, 19
SH	Science Cognitive (Figural Response)	16	16	16	17, 19, 20
SI	Science Cognitive	13	13	4	18, 20, 21

Three additional reading booklets were created for age 9/grade 4 students to bridge the results from the 1990 assessment, which used a separate answer sheet for responses, to the results from the 1988 assessment, where the age 9/grade 4 students responded in the booklet. Each booklet contained a different combination of two reading blocks from the 1990 assessment (RC, RD, or RE) and one block that was used in the 1988 assessment (RJ, RK, or RL), preceded by the student demographics and content area questionnaires from the 1988 assessment (blocks CB and RA, respectively).

Timing

Students were allowed 15 minutes of assessment time for each cognitive block in a booklet, and five minutes of assessment time for each content area questionnaire. With the exception of the age 9/grade 4 students, students were given five minutes to respond to the student demographic questionnaire. To improve the validity of the younger students' answers, the assessment administrators read each demographic question aloud. At the other two age/grade levels, only the (first) race/ethnicity question was read aloud; students read and answered the remaining questions on their own.

4.2 Student Booklets—Bridge Assessments

There were several bridge samples in the 1990 assessment (see Chapter 1), each of which required the use of special booklets. Tables 4-7, 4-8, and 4-9 summarize the contents of each bridge assessment booklet and show how many of each booklet were administered. Tables 4-10, 4-11, and 4-12 give details of the item blocks used in the bridge assessments, including the number of cognitive and open-ended items in each block and the booklets in which each block appeared.

Bridge to 1984. Six booklets (numbered 51 to 56) containing reading and writing items were administered to each age class. These booklets were identical to booklets used in the 1984 assessment of reading and writing and were BIB-spiraled for administration. Each booklet consisted of a common background block (BZ) and three cognitive blocks (at least one reading block and at least one writing block). In addition to cognitive items, the cognitive blocks also contained subject-related background questions.

Bridges to 1986. Three booklets (91, 92, and 93) containing reading, mathematics, and science items were administered to ages 9 and 13. These booklets were identical to those used in the 1986 assessment to measure trends. Each booklet contained a common background block (C1) and three cognitive blocks—one reading block (R1, R2, or R3), one mathematics block (M1, M2, or M3) and one science block (S1, S2, or S3). Mathematics block M3 contained items that required the use of a calculator. All cognitive blocks also contained subject-related background questions.

Table 4-7

Bridge Sample Booklet Contents and Number of Booklets Administered
Age Class 1

Subject Area	Booklet Number	Answer Document[1]	Common Background Block	Subject Area Background Block	Cognitive Blocks			Number of Booklets Administered
BRIDGE TO 1984, AGE 9/GRADE 4								
Reading and Writing	51	C	BZ	—[2]	BC	BL	BQ	978
	52	C	BZ	—[2]	BH	BE	BR	999
	53	C	BZ	—[2]	BC	BK	BJ	1,003
	54	C	BZ	—[2]	BG	BO	BE	982
	55	C	BZ	—[2]	BM	BG	BN	982
	56	C	BZ	—[2]	BV	BR		982
BRIDGE TO 1986, AGE 9								
Reading, Mathematics, and Science	91	B	C1	—[2]	R1	M1	S1	1,991
	92	B	C1	—[2]	S2	R2	M3[3]	2,050
	93	B	C1	—[2]	M2	S3	R3	2,194
BRIDGE FOR LONG-TERM TREND, AGE 9								
Mathematics and Science	94	C	CA	MB	MM	MN	SJ	2,052
	95	C	CA	SB	SK	SL	MO	2,082

[1] B = scannable booklet; C = circled-answer booklet
[2] Subject area background questions included in cognitive blocks for this booklet
[3] Calculator needed for this block

Table 4-8

Bridge Sample Booklet Contents and Number of Booklets Administered
Age Class 2

Subject Area	Booklet Number	Answer Document[1]	Common Background Block	Subject Area Background Block	Cognitive Blocks			Number of Booklets Administered
BRIDGE TO 1984, AGE 13/GRADE 8								
Reading and Writing	51	C	BZ	—[2]	BM	BK	BD	1,044
	52	C	BZ	—[2]	BC	BL	BQ	1,058
	53	C	BZ	—[2]	BH	BE	BR	1,038
	54	C	BZ	—[2]	BN	BC	BD	1,037
	55	C	BZ	—[2]	BG	BO	BE	1,026
	56	C	BZ	—[2]	BG	BJ	BP	1,030
BRIDGE TO 1986, AGE 13								
Reading, Mathematics, and Science	91	B	C1	—[2]	R1	M1	S1	2,229
	92	B	C1	—[2]	S2	R2	M3[3]	2,288
	93	B	C1	—[2]	M2	S3	R3	2,132
BRIDGE FOR LONG-TERM TREND, AGE 13								
Mathematics and Science	94	C	CA	MB	MM	MN	SJ	2,250
	95	C	CA	SB	SK	SL	MO	2,205

[1] B = scannable booklet; C = circled-answer booklet
[2] Subject area background questions included in cognitive blocks for this booklet
[3] Calculator needed for this block

Table 4-9

Bridge Sample Booklet Contents and Number of Booklets Administered
Age Class 3

Subject Area	Booklet Number	Answer Document[1]	Common Background Block	Subject Area Background Block	Cognitive Blocks			Number of Booklets Administered
BRIDGE TO 1984, AGE 17/GRADE 11								
Reading and Writing	51	C	BZ	—[2]	BM	BK	BD	931
	52	C	BZ	—[2]	BC	BL	BQ	939
	53	C	BZ	—[2]	BH	BE	BR	944
	54	C	BZ	—[2]	BN	BC	BD	925
	55	C	BZ	—[2]	BG	BO	BE	936
	56	C	BZ	—[2]	BG	BJ	BP	939
BRIDGE TO 1986, AGE 17/GRADE 11								
Reading, Mathematics, and Science	61	B	C1	—[2]	R2	M4	M2	1,387
	62	B	C1	—[2]	M1	R2	M9[3]	1,359
	63	B	C1	—[2]	S1	S11	R1	1,408
	64	B	C1	—[2]	S2	S4	R5	1,401
	65	B	C1	—[2]	S3	R6	M3[3]	1,375
	66	B	C1	—[2]	R3	R4	R2	1,408
BRIDGE TO 1986, AGE 17								
Mathematics and Science	84	B	C1	—[2]	M1	M2	S3	2,205
	85	B	C1	—[2]	S1	S2	M3[3]	2,206
BRIDGE FOR LONG-TERM TREND, AGE 17								
Mathematics and Science	94	C	CA	MB	MM	MN	SJ	2,193
	95	C	CA	SB	SK	SL	MO	2,209

[1] B = scannable booklet; C = circled-answer booklet
[2] Subject area background questions included in cognitive blocks for this booklet
[3] Calculator needed for this block

Table 4-10

Bridge Sample Block Information, Age Class 1

Block	Type	Total Number of Items	Number of Cognitive Items	Number of Open-ended Items: Cognitive	Number of Open-ended Items: Noncognitive	Booklets Containing Block
BZ	Common Background	37	0	0	1	51 - 56
C1	Common Background	28	0	0	0	91 - 93
CA	Common Background	18	0	0	0	94, 95
BC	Writing Background/Cognitive	23	1	1	0	51, 53
BE	Writing Background/Cognitive	11	2	2	0	52, 54
BG	Writing Background/Cognitive	8	2	2	0	54, 55
BH	Reading Background/Cognitive	15	11	1	0	52
BJ	Reading Background/Cognitive	24	13	1	0	53
BK	Reading Background/Cognitive	19	11	0	0	53
BL	Reading Background/Cognitive	26	7	1	1	51
BM	Reading Background/Cognitive	16	12	1	0	55
BN	Reading Background/Cognitive	25	14	1	0	55
BO	Reading Background/Cognitive	22	11	0	0	54
BQ	Reading Background/Cognitive	21	12	0	0	51
BR	Reading Background/Cognitive	16	12	0	0	52, 56
BV	Reading and Writing Background/Cognitive	36	7 Rd. 1 Wr.	1 Rd. 1 Wr.	0	56
R1	Reading Background/Cognitive	20	9	0	0	91
R2	Reading Background/Cognitive	20	11	0	0	92
R3	Reading Background/Cognitive	17	10	1	0	93
M1	Mathematics Background/Cognitive	26	26	9	0	91
M2	Mathematics Background/Cognitive	26	26	9	0	93
M3	Mathematics Background/Cognitive (Calc.)	19	16	10	0	92
MB	Mathematics Background	14	0	0	0	94
MM	Mathematics Cognitive	20	20	0	0	94
MN	Mathematics Cognitive	13	13	0	0	94
MO	Mathematics Cognitive	16	16	7	0	95
S1	Science Background/Cognitive	23	18	0	0	91
S2	Science Background/Cognitive	25	25	0	0	92
S3	Science Background/Cognitive	31	20	0	0	93
SB	Science Background	15	0	0	0	94
SJ	Science Cognitive	23	23	0	0	94
SK	Science Cognitive	23	23	0	0	95
SL	Science Cognitive	20	20	0	0	95

Table 4-11

Bridge Sample Block Information, Age Class 2

Block	Type	Total Number of Items	Number of Cognitive Items	Number of Open-ended Items: Cognitive	Number of Open-ended Items: Noncognitive	Booklets Containing Block
BZ	Common Background	37	0	0	1	51 - 56
C1	Common Background	30	0	0	0	91 - 93
CA	Common Background	21	0	0	0	94, 95
BC	Writing Background/Cognitive	23	1	1	0	52, 54
BD	Writing Background/Cognitive	25	1	1	0	51, 54
BE	Writing Background/Cognitive	11	2	2	0	53, 55
BG	Writing Background/Cognitive	8	2	2	0	55, 56
BH	Reading Background/Cognitive	18	13	1	1	53
BJ	Reading Background/Cognitive	24	14	2	0	56
BK	Reading Background/Cognitive	17	9	1	0	51
BL	Reading Background/Cognitive	27	6	1	1	52
BM	Reading Background/Cognitive	16	12	1	0	51
BN	Reading Background/Cognitive	23	12	1	0	54
BO	Reading Background/Cognitive	21	10	2	0	55
BP	Reading Background/Cognitive	15	9	1	0	55
BQ	Reading Background/Cognitive	23	17	0	0	52
BR	Reading Background/Cognitive	19	15	0	0	53
R1	Reading Background/Cognitive	31	12	1	0	91
R2	Reading Background/Cognitive	19	10	0	0	92
R3	Reading Background/Cognitive	28	13	0	0	93
M1	Mathematics Background/Cognitive	51	37	9	0	91
M2	Mathematics Background/Cognitive	44	37	8	0	93
M3	Mathematics Background/Cognitive (Calc.)	32	24	10	0	92
MB	Mathematics Background	22	0	0	0	94
MM	Mathematics Cognitive	20	20	0	0	94
MN	Mathematics Cognitive	21	21	0	0	94
MO	Mathematics Cognitive	20	20	0	0	95
S1	Science Background/Cognitive	36	25	0	0	91
S2	Science Background/Cognitive	40	27	0	0	92
S3	Science Background/Cognitive	36	27	0	0	93
SB	Science Background	25	0	0	0	95
SJ	Science Cognitive	24	24	0	0	94
SK	Science Cognitive	23	23	0	0	95
SL	Science Cognitive	30	30	0	0	95

Table 4-12

Bridge Sample Block Information, Age Class 3

Block	Type	Total Number of Items	Number of Cognitive Items	Number of Open-ended Items: Cognitive	Number of Open-ended Items: Noncognitive	Booklets Containing Block
BZ	Common Background	48	0	0	1	51 - 56
C1	Common Background	48	0	0	0	61 - 66, 84, 85
CA	Common Background	30	0	0	0	94, 95
BC	Writing Background/Cognitive	23	1	1	0	52, 54
BD	Writing Background/Cognitive	25	1	1	0	51, 54
BE	Writing Background/Cognitive	11	2	2	0	53, 55
BG	Writing Background/Cognitive	8	2	2	0	55, 56
BH	Reading Background/Cognitive	19	13	1	2	53
BJ	Reading Background/Cognitive	17	6	2	1	56
BK	Reading Background/Cognitive	17	9	1	0	51
BL	Reading Background/Cognitive	32	6	1	2	52
BM	Reading Background/Cognitive	16	12	1	0	51
BN	Reading Background/Cognitive	32	12	1	1	54
BO	Reading Background/Cognitive	24	13	1	0	55
BP	Reading Background/Cognitive	25	11	1	0	56
BQ	Reading Background/Cognitive	17	11	1	0	52
BR	Reading Background/Cognitive	20	9	0	0	53
R1	Reading Background/Cognitive	31	12	1	0	63
R2	Reading Background/Cognitive	19	10	0	0	61, 62, 66
R3	Reading Background/Cognitive	28	13	0	0	66
R4	Reading Background/Cognitive	21	14	0	0	66
R5	Reading Background/Cognitive	18	12	0	0	64
R6	Reading Background/Cognitive	18	14	5	1	65
M1	Mathematics Background/Cognitive	49	35	10	0	62, 84
M2	Mathematics Background/Cognitive	49	35	5	0	61, 84
M3	Mathematics Background/Cognitive (Calc.)	35	24	14	0	65, 85
M4	Mathematics Background/Cognitive	43	29	12	0	61
M9	Mathematics Background/Cognitive (Calc.)	61	41	13	0	62
MB	Mathematics Background	34	0	0	0	94
MM	Mathematics Cognitive	21	21	0	0	94
MN	Mathematics Cognitive	23	23	3	0	94
MO	Mathematics Cognitive	23	23	0	0	95
S1	Science Background/Cognitive	38	27	0	0	63, 85
S2	Science Background/Cognitive	41	32	0	0	64, 85
S3	Science Background/Cognitive	32	23	0	0	65, 84
S4	Science Background/Cognitive	31	20	1	0	64
S11	Science Background/Cognitive	29	20	1	0	63
SB	Science Background	32	0	0	0	95
SJ	Science Cognitive	24	24	0	0	94
SK	Science Cognitive	24	24	0	0	95
SL	Science Cognitive	30	30	0	0	95

Six booklets (61 to 66) containing reading, mathematics, and science items were administered to age 17/grade 11 students. These booklets were identical to booklets used in the 1986 main assessment of reading, mathematics, and science and were BIB-spiraled for administration. Each booklet contained a common background block (C1) and three cognitive blocks. Six reading blocks (R1 - R6), five mathematics blocks (M1 - M4, M9), and five science blocks (S1 - S4, S11) were assembled in various combinations to create the booklets. Mathematics blocks M3 and M9 contained items that required the use of a calculator. All cognitive blocks also contained subject-related background questions.

Booklets 84 and 85, administered to age 17 students, contained mathematics and science items. These booklets were identical to booklets used in the 1986 assessment and were administered with audiotape pacing. Each booklet contained a common background block (C1) and three blocks of cognitive items—at least one mathematics block (M1 - M3) and at least one science block (S1 - S3). Mathematics block M3 contained items that required the use of a calculator. All cognitive blocks also contained subject-related background questions.

Bridge for Long-term Trend. Booklets 94 and 95, administered to ages 9, 13, and 17, contained mathematics and science items that had last been administered in an assessment prior to 1986. Booklet 94 contained a common background block (CA), a mathematics background block (MB), two blocks of mathematics items (MM and MN) and a block of science items (SJ). Booklet 95 contained the common background block, a science background block (SB), two blocks of science items (SK and SL) and a block of mathematics items (MO).

4.3 Excluded Student Questionnaire

Some students selected for the assessment were judged by school authorities to be incapable of participating in the assessment because they had limited English language proficiency, were mildly mentally retarded (educable), or were functionally disabled.

The exclusion criteria for the 1990 main assessments differed from those used for the 1990 bridge assessments and for assessments prior to 1990. To be excluded from the main assessment, sampled students had to be identified by the school as Limited English Proficient or having an Individualized Education Plan *and* (in either case) judged incapable of participating in the assessment.

For each student excluded from the assessment, schools were required to complete a questionnaire containing 27 questions about the characteristics of that student and the reason for exclusion. For students with an Individual Education Plan, the questionnaire included questions about students' functional grade level, mainstreaming, and special education programs. For Limited English Proficient students, it asked about students' native language, time spent in special education and language programs, and the level of the students' English language proficiency.

Of the 90,531 students sampled for the main assessments, a total of 5,728 (6.3 percent overall) were excluded from the assessments: 2,332 (7.2 percent) at age class 1, 1,950 (6.7 percent) at age class 2, and 1,446 (5.1 percent) at age class 3.

Of the 56,397 students sampled for the bridge assessments, 3,450 (6.1 percent overall) were excluded from the assessments: 1,116 (6.8 percent) at age class 1, 1,095 (6.3 percent) at age class 2, and 1,239 (5.4 percent) at age class 3.

4.4 Teacher Questionnaires

To supplement the information on instruction reported by students, the teachers of fourth and eighth graders participating in the mathematics assessments were asked to complete a teacher questionnaire about their instructional practices, teaching backgrounds, and characteristics. In addition, administration of an eighth-grade science teacher questionnaire was made possible by supplemental funding from the National Science Foundation through a subcontract from Horizon Research, Inc. to ETS.

The first part of the teacher questionnaires pertained to the teachers' background and training and included questions (34 for mathematics; 100 for science) pertaining to gender, race/ethnicity, years of teaching experience, certification, degrees, major and minor, coursework in education, coursework in subject area, in-service training, extent of control over classroom, instruction, and curriculum, and availability of resources for classroom.

The second part of the questionnaires pertained to the procedures used for *each class* containing an assessed student and included questions (at grade 4, 34 questions; at grade 8, 35 questions for mathematics and 58 questions for science) on the ability level of students in the class, whether students were assigned to the class by ability level, time on task, homework assignments, frequency of instructional activities used in class, instructional emphasis given to the topics and skills covered in the assessment, and use of particular resources.

Responses were received from 882 fourth-grade mathematics teachers, 597 eighth-grade mathematics teachers, and 510 eighth-grade science teachers.

4.5 School Characteristics and Policies Questionnaire

A school characteristics and policy questionnaire was given to the principal or other administrator of each school that participated in the 1990 NAEP assessment. This questionnaire asked questions (117 at age classes 1 and 2, 125 at age class 3) about background and characteristics of school principals, length of school day and year, school enrollment, absenteeism, drop-out rates, size and composition of teaching staff, policies about tracking, curriculum, testing practices and use, special priorities and school-wide programs, availability of resources, special services, community services, policies for parental involvement, and school-wide problems.

Responses were received from 753 of the 815 participating schools at age class 1, 635 of the 688 participating schools at age class 2, and 527 of the 596 participating schools at age class 3.

4.6 Principal Questionnaire

Before the assessment, Westat, Inc., distributed a questionnaire to the principal of each participating school to gather data about school characteristics, including school enrollment and attendance, parents' occupations, and student race/ethnicity. These data were used in part to estimate the number of age/grade-eligible students and to determine the correct "size and type of community" classification for each school.

4.7 "The Nation's Portfolio": The 1990 NAEP Study of School-based Writing

As part of the 1990 writing trend assessment, NAEP conducted a special study of school-based writing, involving random samples of approximately 2,000 students at age 9/grade 4 and another 2,000 students at age 13/grade 8. The purpose of this "portfolio study" was to expand the view of students' writing abilities by analyzing pieces of writing that they produced outside of the assessment situation.

NAEP wrote to the English/language arts teachers of students sampled for the study and asked these teachers to work with the students to select an example of their best writing, prepared in response to an assignment for English or language arts class. The NAEP letter expressed an interest in collecting student papers that had been developed using a writing process approach—that is, papers in which students prepared and revised successive drafts and shared the writing with others. Finally, to provide a context for analyzing the portfolio contents, the teachers were asked to attach a copy or description of the assignment for which the paper was written, together with a brief description of any associated writing activities (for example, number of drafts, use of peer or teacher review, or nature of revisions). These materials were mailed to NCS and forwarded to ETS for analysis and reporting.

At age 9/grade 4, the response rate was 54 percent; papers were received for 1,066 of the 1,962 students who were sampled for the school-based writing study. At age 13/grade 8, the response rate was 51 percent; papers were received for 1,059 of the 2,071 sampled students. The rather low participation rates were due in part to the method of collecting the portfolio data. Teachers were asked to mail in students' samples of writing, rather than handing them in to NAEP field administrators.

Chapter 5

FIELD OPERATIONS AND DATA COLLECTION

Nancy Caldwell, Renee Slobasky, Dward Moore, and James Ter Maat

Westat, Inc.

As a subcontractor to Educational Testing Service, Westat, Inc., was responsible for field operations and data collection for the 1990 assessment. This chapter summarizes these activities; details are provided in the *Report on Field Operations and Data Collection Activities, 1990 National Assessment of Educational Progress* (Caldwell, Slobasky, Moore, & Ter Maat, 1991).

In 1990, NAEP involved two components: the national assessment and the Trial State Assessment. The national part of the 1990 NAEP was conducted in 38 states, including the District of Columbia; a total of 40 states and other jurisdictions volunteered for the 1990 Trial State Assessment. With so many jurisdictions involved in each of the two NAEP components, there was substantial overlap (Table 5-1). In fact, 30 jurisdictions participated in both national and Trial State assessments. To the extent possible, the school samples were designed so that no school would be selected for both NAEP components. However, because of sampling requirements, 14 very large schools were selected for both.

Westat's involvement in the design and conduct of data collection operations for the 1990 Trial State Assessment began in 1988, with the preparations for the 1989 field test. It continued until August 1990, with the delivery of data collection summary reports to those states participating in the 1990 effort. The details of these activities for the Trial State Assessment are given in the *Report on Field Operations and Data Collection Activities*.

The following sections describe the field operations and data collection effort for the national portion of the 1990 assessment.

5.1 Overview of Field Operations and Data Collection for the 1990 National Assessment

The 1990 national assessment was conducted in a sample of approximately 2,400 public and private schools located in 108 geographic areas called primary sampling units (PSUs) in 38 states including the District of Columbia. The 108 areas were selected by Westat to represent the nation as a whole. Assessments were conducted throughout the school year, although the majority occurred from January to May 1990.

In order to reduce the burden on the participating schools, national assessment field staff did most of the work associated with the assessment. Introductory meetings were h~ld in the fall to explain the assessment procedures to school and district representatives and to set a

Table 5-1

Participants in the Two Components of the 1990 NAEP

Participants in Both National and Trial State Assessments	
Alabama	Maryland
Arizona	Michigan
Arkansas	Minnesota
California	Montana
Colorado	New Jersey
Connecticut	New Mexico
Delaware	New York
District of Columbia	North Carolina
Florida	Ohio
Georgia	Oklahoma
Hawaii	Oregon
Illinois	Pennsylvania
Indiana	Texas
Iowa	Virginia
Louisiana	Wisconsin

Participants in the National Assessment Only	
Maine	South Carolina
Massachusetts	Tennessee
Mississippi	Utah
Missouri	Washington

Participants in the Trial State Assessment Only	
Guam	Rhode Island
Idaho	Virgin Islands
Kentucky	West Virginia
Nebraska	Wyoming
New Hampshire	

mutually agreed-upon assessment date. The assessment supervisor visited the school to select the sample of students a week or two before the assessment. The assessment sessions were conducted by national assessment field staff (the exercise administrators) under the direction of the assessment supervisors. At the completion of the assessment in a school, field staff coded the booklets and shipped the completed materials to NCS, the processing subcontractor.

5.1.1 Field Organization

The field operations and data collection efforts were organized into three time periods: fall, winter, and spring. During the fall time period (mid-September to mid-December 1989), NAEP/Westat field staff contacted districts for cooperation, conducted introductory meeting, hired exercise administrators, and conducted the fall assessment. During the winter period (January to mid-March 1990), the winter assessments were carried out. The spring assessments were conducted during the spring period (mid-March to mid-May 1990).

Under the direction of the Westat home office staff, 9 area supervisors led the activities of 38 assessment supervisors, who oversaw the work of the exercise administrators (1 per PSU during the fall period, and 1 to 2 per PSU in the winter and spring periods). During the fall period, 31 state supervisors also participated in certain aspects of the national portion of the assessment field operation under the direction of the area supervisors.

5.1.2 Schedule of Project Activities

Summarized in Table 5-2 are all of the scheduled project activities for the 1990 national assessment. Details are provided in subsequent sections of this chapter.

Table 5-2
Summary of the Field Operations and Data Collection Schedule

Date	Activity
Early June 1989	Department of Education sent first letter to Chief State School Officers (CSSOs) informing them that schools within their states had been chosen for one or both components of NAEP.
Mid-June 1989	ETS sent followup letter to Chiefs with lists of school districts and numbers of schools selected for NAEP.
Early August 1989	ETS sent all district superintendents and private schools a letter about NAEP and a set of NAEP reports.
Mid-August 1989	Westat sent all districts a list of schools selected for either or both programs and a list of the sample frame to be checked for omissions.
Mid-September 1989	Westat held first supervisor training session.

Date	Activity
Late September to mid-December 1989	Supervisors contacted district superintendents and private school principals about the national assessment, to establish or confirm participation and to make arrangements for introductory meetings with school representatives. Westat sent letter to superintendents confirming the date, time, and place of introductory meeting. Package of materials sent to principals of participating schools. Supervisors conducted introductory meetings for the national assessment.
Mid-October 1989	ETS sent all participating schools a brochure and a set of NAEP reports.
October 9 - December 13, 1989	Fall bridge assessments
Mid-December 1989	Second supervisors training session.
Early January 1990	Supervisors prepared for winter assessment.
January 8 - March 16, 1990	Winter assessments
March 19 - May 18, 1990	Spring assessments

5.2 Pre-assessment Activities

During the fall period (mid-September through mid-December 1989), a number of activities were conducted for the national assessment, including telephone contacts to district superintendents and private school principals to gain their participation, conducting introductory meetings with school principals to explain the national assessment, and conducting the fall assessments in about 290 schools (from October 9 to December 13, 1989).

5.2.1 Supervisor Training

The 80 assessment supervisors and state supervisors came to Bethesda, Maryland for a five-day training session from September 15-19, 1989. Also in attendance were representatives from Educational Testing Service (ETS), National Computer Systems (NCS), and the National Center for Education Statistics (NCES). The training was conducted by the Westat project director and field director assisted by the field managers. ETS Princeton office staff also made presentations and provided explanatory notes throughout the session.

The training session focused on components of both the national assessment and the Trial State Assessment. While the state supervisors were trained separately in preparation for the Trial State Assessment, the entire group of supervisors were trained on the general features of the Trial State Assessment, anticipating that the assessment supervisors would be questioned about the Trial State Assessment by district and/or school personnel. Also, there were presentations to all supervisors on the national assessment since the state supervisors would be conducting some introductory meetings for the national assessment.

Topics on the training agenda included an overview of the national assessment and the Trial State Assessment and the supervisors' responsibilities, a discussion of various reports from recent assessments, procedures for contacting districts and conducting introductory meetings, scheduling assessments within the PSUs, recruiting and training exercise administrators, procedures for drawing the sample of students, conducting assessments, preparation and distribution of questionnaires, and administrative forms and procedures. Also featured were practice exercises in sampling and filling out the various administrative forms.

In response to requests from supervisors in their evaluations of the 1986 and 1988 assessments, a mock assessment session was held with the supervisors acting as "students." This included reading verbatim from one of the actual scripts to be used during an assessment and following prescribed procedures for distributing materials, reading directions, and recording the results of the assessment.

5.2.2 Gaining Cooperation of Sampled Schools

The process of gaining cooperation of the schools selected for the national assessment began in the summer, 1989, with a series of letters and contacts with state and district level officials. The schedule of these contacts is presented in Table 5-2.

Recruiting of schools for NAEP actually began in June, once the sample of schools had been selected and their corresponding school districts identified. NCES and ETS contacted the Chief State School Officers in each state notifying them of the districts and schools in their states that were in the sample. In the 30 states participating in the Trial State Assessment that had schools sampled for the national assessment, the Trial State Assessment state coordinator was also sent the list of districts and schools sampled for the national assessment.

In August, ETS sent a set of recent NAEP reports and a letter to the superintendents and heads of private schools inviting their participation. In mid-August, Westat sent districts two lists—a list of the sampled schools and a list of the schools in the sample frame to be checked for omissions. These initial contacts, which were completed prior to supervisor training, paved the way for the telephone contacts to follow.

Once the supervisors had been trained, they began working on obtaining cooperation. The contacting of districts proceeded differently in states participating in the Trial State Assessment than in the states that were not. In states participating in the Trial State Assessment, the state supervisor first spoke with the state coordinator to determine what, if any, contacts had been made with districts about the national assessment. The state supervisor then notified the assessment supervisor about the extent of contact that the state coordinator had had with these districts. Contacts began immediately following training with private schools, parochial districts, and with districts in the eight states that were participating only in the national assessment.

As the supervisors contacted superintendents and private school officials to establish cooperation and to set up the introductory meetings, they completed two forms. The Introductory Meeting Form was used to record the names of the schools and individuals expected to attend each meeting. A Results of Contact form was completed documenting the

discussion the supervisor had with each administrator concerning the district's willingness to participate and any special circumstances regarding the introductory meeting or assessments.

Copies of these forms were sent to the area supervisor and to the home office. Once received in the home office, the forms were used as the basis for mailing packages of materials to the persons scheduled to attend the meeting.

As indicated by Table 5-3, school cooperation rates have remained above 85 percent since the NAEP 1984 assessment.

Table 5-3
School Response Rates, NAEP 1984 - 1990

	1984	1986	1988	1990
Rate	88%	87%	87%	86%
Total cooperating	1,361	1,633	1,412	1,970

5.2.3 Introductory Meetings

During the period from late September through the middle of December, supervisors visited all 108 PSUs in the national sample and conducted introductory meetings. The supervisors had a number of tasks to perform during the introductory meetings. While the content of the meetings varied, they generally included:

- collecting and checking completed principal questionnaires;
- presenting an overview of NAEP;
- answering questions;
- explaining the tasks required of each school;
- setting preliminary sampling and assessment dates for each school;
- verifying information on and completing the School Control Form;
- distributing and explaining the Student Listing Forms;
- identifying a school coordinator; and
- inquiring about possible exercise administrator candidates.

Introductory meetings generally lasted about one hour. They ranged in size from small meetings between the supervisor and one school coordinator to formal meetings attended by 20 to 30 school officials (superintendents, curriculum specialists, testing personnel, principals, and coordinators). The introductory meetings were the first opportunity for principals and other officials at the school level to discuss the assessment with NAEP staff. Thus, the meetings were particularly important for establishing rapport with the schools, assuring school cooperation, and explaining the details of the schools' tasks to the individuals responsible for them.

5.2.4 Making Arrangements for the Assessments

During the introductory meetings, the supervisor discussed arrangements for the assessments with representatives from each school. Within the weeks scheduled for each PSU, the supervisor had the flexibility to set each school's assessment date in coordination with school staff. The staff sometimes expressed preferences for a particular day or dates or had particular times when the assessment could not be scheduled. Using this information, the supervisors set up the assessment schedule for the PSU.

The School Control Form was used by the supervisors to record information about the school's assessment plan. The form gave estimates of the number of students to be assessed in the school as well as the type of sessions to be held.[1] Using this information, the supervisor and school staff could discuss the approximate number of sessions to be held in the school and the space required.

The supervisor usually learned during the introductory meeting whether a school required some form of parental notification or permission. In preparation for this, the supervisor had copies of three versions of standard NAEP letters to parents. These letters were made available to schools requesting them. If they preferred, schools could send out their own letters and notices.

Following an introductory meeting, the supervisor sent the principal questionnaires and copies of the School Control Forms to the home office. State supervisors who were conducting meetings for assessment supervisors sent them their copy of the School Control Form.

5.2.5 Recruiting, Hiring, and Training Exercise Administrators

During the fall, while the supervisors were conducting introductory meetings and scheduling assessments, they also recruited and hired exercise administrators. The exercise administrator's primary job was to administer the assessment sessions. Exercise administrators were recruited from many sources. Each supervisor was given a computer list of interviewers and exercise administrators who had worked for Westat on other studies, including the 1988 NAEP. During introductory meetings, the supervisors asked the school principals and other

[1] Schools in the main NAEP samples could have up to three different session types; schools in the bridge samples could have up to four different session types (see Chapter 3 for definitions of session types for the various samples).

staff to recommend potential exercise administrators. Where necessary, advertisements were placed in local newspapers and the job service was notified.

Supervisors were told that, in general, two exercise administrators should be hired for each PSU, although a variety of factors might influence the actual number. The number of schools in a PSU, the size of the student sample in each school, distances to be traveled, the geography of the area, and weather conditions during particular times of the year were all factors taken into consideration by supervisors in developing their plan for exercise administrators.

A few supervisors who had contiguous PSUs hired the same exercise administrators to work in all their PSUs. Supervisors who had PSUs where schools were small and widely scattered tended to hire exercise administrators to work only a portion of the PSU. Supervisors were encouraged to hire locally and to hire individuals with teaching experience or the ability to handle classroom situations. Many of the approximately 400 NAEP 1990 administrators were retired or substitute teachers.

The assessment supervisors had complete responsibility for recruiting, hiring, training, and supervising their exercise administrators. The supervisors' first task upon arriving in a PSU for the assessments was to train the administrators. Exercise administrators were required to study the Exercise Administrator's Manual before attending a half-day training session conducted by the supervisor. During the training, the supervisor reviewed, in detail, all aspects of the administrator's job including preparing materials, booklets, and Administration Schedules for assessments; the actual conduct of the session; post-assessment collection of booklets and other assessment materials; coding booklet covers; recordkeeping; and administrative matters.

5.3 The Fall Bridge Assessment

To provide continuity and comparability with the past, the fall bridge assessment replicated what had been done in prior years for age 13/grade 8. Tape sessions were conducted with samples of age-eligible students, as had been done in all previous years. Additional samples of age- and grade-eligible students were assessed with spiral (self-administered) booklets, following procedures initiated in NAEP 1984.

Each of the 38 assessment supervisors was responsible for conducting the fall assessments in an average of seven schools each. The nine-week fall assessment time period required the supervisors to sample and assess from one to two schools per week. On average, there were two to three schools per PSU.

5.3.1 Selecting the Student Sample

Two weeks prior to a school's assessment date, the assessment supervisor contacted the school coordinator to make sure that the lists of eligible students were prepared and that all arrangements were set as agreed. The supervisor then visited the school (or district office) a few days to a week or more before the assessment date to select the sample of students.

The supervisor's first task upon arriving at the school to select the sample was to review the Student Listing Forms or comparable list of students in an effort to be sure that they had been completed correctly. The supervisor made certain checks to help assure that all age- and all grade-eligible students had been listed. The supervisor also checked that the students to be excluded from the assessment were listed so that they could be included in the sample information.

For each school, the Westat home office produced a Session Assignment Form, which told the supervisor how to select the sample in that school. The Session Assignment Forms contained the following information:

- *Identifying Information* - identifying the school, the age and grade level, and the approximate number of students to be assessed in total and by session type.

- *Type of Sessions* - specifying whether the school was to have spiral only, tape only, or both spiral and tape sessions. If tape sessions were assigned to the school, then the types of tape sessions were specified.

- *Sampling Instructions* - the particular steps the supervisor should follow in selecting the sample of eligible students in the school:

  - Review the lists for completeness.

  - Consecutively number all eligible students.

  - Compare the number of eligibles to the minimum and maximum specified on the Session Assessment Form. If the actual number of eligibles was out of range, the supervisor had to call Westat for sampling advice.

  - Select the sample of students as specified on the Session Assignment Form. The steps to be followed depended on whether the school was selected for spiral, tape, or both. If the school was selected for tape (for which only age-eligible students are considered), the supervisor was to mark the age-eligible students after marking the selected line numbers. The eligible students were then assigned to the appropriate tape sessions, if there was more than one.

Following the sampling instructions, the supervisor was instructed to fill out an Administration Schedule for each session listing the sampled students. Before listing the students on the Administration Schedules, the supervisor reviewed the plans for the assessment with the school coordinator. If, for example, a large number of students were sampled for a spiral session, the supervisor discussed Westat's preference for this group to be divided into sessions of about 30 each.

After any excluded students were identified, the supervisors were instructed to prepare and distribute an excluded student questionnaires for each excluded student. If the coordinator could not identify the excluded students while the supervisor was at the school, a set of instructions for excluding students was left with the coordinator along with an estimated number of questionnaires needed.

For spiral sessions, the final step on the Session Assignment Form was for the supervisor to prepare a list of names of English teachers for use in the writing portfolio study (see section 5.3.2.1).

5.3.2 Conduct of the Assessment

The primary responsibility for conducting assessment sessions was held by the exercise administrators. Supervisors were required to observe the first session an exercise administrator conducted to ensure that he or she followed the procedures properly. Supervisors were also required to be present in all schools during the assessments if at all possible, especially in large schools with several sessions.

To ensure that sessions were administered in a uniform way, the exercise administrator was provided with scripts for each session type from which he or she was to read verbatim. The scripts began with a brief introduction to the study followed by directions to the administrator to distribute the booklets, being careful to give each student the correct preassigned booklet.

Following the distribution of booklets, the scripts differed depending on whether the session was a spiral or tape session. In spiral sessions, the exercise administrator read from the script and followed its directions as he or she continued the session administration and timed the sections of the booklets. In tape sessions, the administrator was instructed to turn the tape recorder on after distributing the booklets and the tape did most of the administration and timing of the sections.

During the sessions, the exercise administrator monitored the students to be sure that they were working in the correct section of the booklet and to discourage them from looking at a neighbor's booklet. During the background (first) section, the administrator was allowed to assist students in understanding questions and responding to them. After the students began working on the other sections of the booklets, the administrator was not allowed to answer any questions.

At the end of an assessment session, booklets were collected and the students dismissed according the school's policy. The exercise administrator was then responsible for completing the information at the top of the Administration Schedule and coding the covers of the completed booklets.

For each student absent from the session (and from the makeup session, if one was held) an Absent Student Form was prepared by the exercise administrator. This one-page form, similar to that of the booklet covers, allowed NCS to scan and record demographic information on absent students.

5.3.2.1 The Writing Portfolio Sample

After the assessment was over in schools with spiral assessments, the supervisor reviewed the Administration Schedule to identify students who had received booklets 54 and 55. These students' names were listed on a letter given to their English teachers, requesting a sample of

each student's school writing and a copy of the assignment for which the piece was written. Students' names were removed from their writing samples and replaced with their booklet ID number before the samples were removed from the school. The students' writing samples were sent to NCS, either with the shipment of completed booklets or in a separate mailing. The materials were then forwarded to ETS for evaluation.

5.3.3 Results of the Fall Bridge Assessment

Shown in Table 5-4 are preliminary data (based on the reports sent by supervisors to Westat) on the number of students who were sampled, invited, and assessed during the fall assessment of age 13/grade 8. Final numbers were determined by NCS based on the actual booklets received at their Iowa facility.

Table 5-4
Students Invited and Assessed in the Fall Assessment, by Session Type*

Age 13/Grade 8 Students	Session Type					
	Spiral	Tape 91	Tape 92	Tape 93	Tape 94	Tape 95
Number invited	6,869	2,489	2,568	2,363	2,543	2,434
Number assessed	6,222	2,241	2,288	2,143	2,266	2,217
Percent assessed	91%	90%	89%	91%	89%	91%

* These data are based on field reports sent to Westat and do not reflect receipt of actual assessment materials.

The overall attendance rate (90.2 percent) is slightly better than in the 1988 NAEP (90 percent) and the 1986 NAEP (89.7 percent).

A total of 1,119 students who were sampled for the assessment were excluded from participation by the school because they had limited English language proficiency, were judged as being mildly mentally retarded (educable), or were functionally disabled. The rate of exclusion (5.5 percent) remains comparable with the two previous assessments of 1988 (6 percent) and 1986 (5.4 percent).

5.3.4 Assessment Questionnaires and Reports

Each school in the fall assessment was mailed a school characteristics and policies questionnaire by Westat prior to the assessment. This form was to be filled out by the principal or another staff member knowledgeable about the school's administrative policies and staff characteristics.

An excluded student questionnaire was to be filled out for every student who was sampled for the assessment but excluded by the school. Following guidelines used in previous assessments, schools could exclude students who were of limited English speaking ability, mildly mentally retarded (educable), or functionally disabled, if in the judgment of school staff or if school records indicated they were unable to take the assessment. After the sample of students was drawn and the Administration Schedules prepared, the supervisor requested that the coordinator identify any students who should be excluded. The supervisor then gave an excluded student questionnaire to the coordinator to complete for every excluded student.

The supervisor attempted to collect all completed school characteristics and policies and excluded student questionnaires on the assessment day. If the questionnaires were not ready, and it was convenient for the supervisor or an exercise administrator to return to the school later to pick up the questionnaires, they would do so. Otherwise, the supervisor gave the coordinator a postage-paid envelope to be used to mail the forms to NCS.

Once the assessments were finished in a school, the supervisor and/or exercise administrators edited the booklets, filled out the necessary forms, and shipped the booklets and forms to NCS. A copy of all forms was sent to Westat so that progress in the field could be monitored.

The School Worksheet was used to summarize the results of the assessment sessions in each school. The numbers of students to be assessed, assessed, and absent were entered so that the supervisor could calculate if a makeup session was required. If a makeup was required for one or more session types, the supervisor discussed the scheduling of the makeup with the coordinator. The top (original) copy of the School Worksheet, Roster of Questionnaires, and the Administration Schedules (with the student names removed and left at the school) were included with the booklets in the shipment to NCS. For tracking and identification of assessment materials, the supervisor included as necessary in each shipment the Shipping Transmittal Form, the Supplemental Transmittal Form, the Return Shipment Notice, and the Session Header Form.

5.4 The Winter/Spring Main NAEP Assessments

The winter and spring assessments were much larger and more complex than the fall assessment. While the fall assessment involved only one age and grade group, winter/spring included main NAEP assessments of age 9/grade 4, age 13/grade 8, and age 17/grade 12 students and bridge assessments of age 9/grade 4 and age 17/grade 11. The total number of students to be assessed went from about 20,000 in the fall to about 150,000 in the winter/spring.

The winter/spring main NAEP assessments involved two types of spiral sessions (reading/science and mathematics) and, for the first time, a tape session for which age and grade were used as eligibility requirements. A school might be selected for spiral, tape, or both spiral types and tape.

This section discusses the student sample selection, supervisor training, and results of the winter/spring main NAEP assessments.

5.4.1 Supervisor Training

The assessment supervisors in charge of the winter/spring assessment activities were trained December 14-16, 1989, in New Orleans, Louisiana. After a review of the fall assessment, training first focused on the winter and spring bridge assessments (sampling and other procedures; see section 5.5), followed by main NAEP assessment activities (sampling, excluding students, teacher survey, and administrative forms and procedures).

At various times during training, the supervisors were divided into nine smaller groups led by the area supervisors for practice exercises in student sample selection and the various administrative forms for which the supervisors were responsible. Westat home office field managers monitored the training in these groups.

5.4.2 Selecting the Student Sample

Upon arriving at the school (or district office) to select the sample, the supervisor first reviewed the lists of eligible students. The supervisor confirmed with the school coordinator that all eligible students had been listed. If any eligible students had been left off the lists, sampling could not proceed until problems were corrected.

Using the school's Session Assignment Form, the supervisor selected the sample of students to be assessed. After making sure that all eligibles had been listed, the supervisor numbered the students. If the total number of eligible students was within the minimum and maximum limits indicated on the Session Assignment Form, the supervisor could proceed to select the sample. If the number was outside the limits, he or she called Westat for sampling instructions. The supervisor then proceeded to select the sample of students as specified on the Session Assignment Form. The forms provided step-by-step instructions for sampling, as they did in the fall.

Once the students had been assigned to sessions, the supervisor and exercise administrators filled out an Administration Schedule for each session. The supervisor discussed the final schedule of the sessions with the coordinator and the date, time, and location of each session was filled in on the Administration Schedules.

The supervisor then reques d the school coordinator to identify any student having an individualized education plan (IEP) and/or of limited English proficiency (LEP) on the Administration Schedules and whether any of these students should be excluded from the assessment based on the criteria for excluding students in main NAEP schools (see section 5.4.4). For each excluded student, an excluded student questionnaire was then prepared by the coordinator. If the coordinator could not identify the excluded students while the supervisor was at the school, the instructions were left with the coordinator along with an adequate number of questionnaires.

5.4.3 Results of the Assessment

Table 5-5 provides preliminary information (as recorded by Westat) on the number of students sampled, invited to the assessment, and assessed during the winter and spring main NAEP. Final numbers were determined by NCS based on the actual booklets received.

Table 5-5
Students Sampled, Invited, and Assessed
Winter and Spring Main Assessments*

Students	Age 9/Grade 4	Age 13/Grade 8	Age 17/Grade 12	Overall
Number sampled	37,413	35,102	36,537	109,052
Number excluded	2,420	2,171	1,549	6,140
Number invited	34,993	32,931	34,988	102,912
Number assessed	32,525	29,322	28,368	90,215
Percent assessed	92.9%	89.0%	81.0%	87.6%

* These data are based on field reports sent to Westat and do not reflect receipt of actual assessment materials.

The rate of participation among age 9/grade 4 students, which has historically been the highest of the three cohorts in the assessment, remained high in the 1990 NAEP (92.9 percent). At age 17/grade 12, the response rate was 81 percent, higher than it has been since the 1984 NAEP, when it was 82.8 percent (for age 17/grade 11).

Of the almost 110,000 students sampled for assessment, 5.6 percent (6,140) were excluded by schools. Of those students invited to assessment, over 87 percent were assessed. Overall, 90,215 students were assessed during the 20 weeks of the winter and spring main NAEP.

5.4.4 Assessment Questionnaires and Reports

As in the fall, Westat sent each school in the winter/spring main NAEP assessment a school characteristics and policies questionnaire a few weeks before the assessment was scheduled to be conducted. Likewise, supervisors prepared an excluded student questionnaire for each sampled student the school deemed incapable of being assessed following the NAEP exclusion criteria. The criteria and procedures used for determining excluded students in the 1990 main assessment samples differed somewhat from those of previous assessments years and were the same as those used for the 1990 Trial State Assessment.

First, the school identified any sampled students who were classified as limited English proficient (LEP) or had an individualized education plan (IEP) for reasons other than being gifted or talented. Next, the decision was made whether any of these students should be excluded. Even with these new, somewhat more specified exclusion criteria, the percent of

excluded students was comparable with that of recent assessment years (5.6 percent in 1990, 5 percent in 1988, 5 percent in 1986).

To gather data about the teachers of assessed students, a teacher questionnaire was given to selected teachers of fourth-grade mathematics, eighth-grade mathematics, and eighth-grade science. The teachers asked to participate were the mathematics or science teachers of students who were administered mathematics or science booklets. After student sampling was complete, the supervisor requested a list of the teachers of all sampled students, except excluded students and students not taking the subject. The teachers' names were recorded on the Roster of Teacher Survey Questionnaires. Then, the teacher code from the Roster for each student's teacher and his or her class period number were recorded on the Administration Schedule.

The supervisor requested that the teacher questionnaires distributed on the day of sampling be returned if possible by the day of the assessment. For those not returned on assessment day and those distributed after the assessment was completed. a postage-paid envelope was left with the coordinator.

When the assessment was completed in a school, an Absent Student Form was completed for each invited student who did not attend a session. The information at the top of the Administration Schedules was completed (number of students assessed and number absent) and the School Worksheet filled in summarizing all assessment information for the school.

5.5 The Winter and Spring Bridge Assessments

The winter and spring bridge assessments replicated procedures and materials that had been used in years prior to NAEP 1990. The bridge assessments for age 9/grade 4 were held in the winter; age 17/grade 11 in the spring. These assessments were conducted by the assessment supervisors at the same time as the winter/spring main NAEP assessments.

5.5.1 The Winter Bridge Assessment

The winter bridge assessment of age 9/grade 4 was held during the ten-week period from January 8 to March 16, 1990. Six different types of booklets were used during this assessment—one spiral and five different tapes. In a participating school, a supervisor might conduct up to four different session types depending on the size of the school.

For each bridge school, a computer-generated Session Assignment Form gave the supervisor specific sampling instructions. The procedures the supervisors followed in sampling and preparing for the winter bridge assessments were the same as those for the other assessments.

Of the 356 sampled schools for the winter bridge, 288 schools cooperated for a response rate of 87 percent. The student participation rate of 92.5 percent remains high and compares favorably with previous years (92.2 percent in 1988 and 92.6 percent in 1986).

5.5.2 The Spring Bridge Assessment

The spring bridge assessment was held during the nine-week period from March 19 to May 18, 1990. The spring bridge replicated procedures used in assessments prior to 1988. Student eligibility was also based on criteria used in years prior to 1988 when the modal grade changed from the 11th to the 12th grade. Thus, the 1990 spring bridge assessment involved age 17/grade 11.

Six different session types were conducted during the spring bridge - two different spiral and four different tapes. Following the sampling instructions on the Session Assignment Form, supervisors randomly assigned students to spiral, tape, or a combination of spiral and tape sessions (up to four different session types in a school).

Of the 397 schools sampled for the spring bridge, 291 schools participated, or 79 percent. The student cooperation rate of 82.1 percent was an increase over the 79.2 percent assessed in 1988.

5.6 Field Management

Several approaches were taken to monitor the progress of field work throughout the 1990 assessment period. During the pre-assessment activities (arrangements for and conduct of introductory meetings), the assessment supervisors reported to the area supervisors at least once a week to review their progress in scheduling introductory meetings and to discuss any problems or difficulties they were having. The area supervisors also reported once a week to the Westat home office field managers.

In addition, an automated management system was designed that contained a record for each sampled school. A disposition code structure was developed to indicate the status of the school's participation. Each area supervisor maintained a computer file of the schools for which his or her assessment supervisors were responsible. When the assessment supervisors reported the results of their contacts with superintendents, the area supervisors keyed a cooperating disposition code for each school. If a school or school district refused, as noted on the Results of Contact form, a refusing disposition code was keyed along with a brief explanation.

The area supervisors transmitted data weekly from their computers to the home office. Disposition reports were then generated from the receipt system once a week in order that home office staff could review the progress of securing cooperation from the sampled schools. These reports were an invaluable tool for the sampling statisticians as well as for the field director and field managers. They provided the statisticians with the information needed to determine whether the sample of schools was adequate to produce representative results; if necessary, the sampling statisticians substituted schools into the sample to replace some of the noncooperating schools.

During the assessment activities, the assessment supervisors maintained their close reporting relationship with the area supervisors. The automated management system was moved to the home office and expanded to include the results of the actual assessments from

forms such as the School Worksheets, the Roster of Questionnaires, and the Roster of Teacher Survey Questionnaires.

A weekly response rate report allowed the project staff to monitor the progress of the assessments both in terms of checking that the schools were assessed on schedule as well as assuring that a high response rate was achieved. The sampling statisticians used these reports to monitor the sample yield by school, PSU, and age/grade level.

Progress of the assessments was constantly monitored through telephone reports held between the area supervisors and assessment supervisors and between the area supervisors and the home office staff. During these phone conversations, the supervisors' schedules were reviewed as well as any problems that the supervisors were experiencing. The assessment supervisors also called the field managers directly if the area supervisor was unavailable and a situation developed needing immediate attention.

The assessment supervisors filled out a Work Schedule giving their schedule for a one- to two-week period so that they could be contacted if necessary. It also allowed the field managers to review the supervisors' schedules and the distribution of work.

Progress of the fieldwork was also monitored during quality control visits made to the field by Westat and ETS home office staff. The results of the quality control visits are given in the Westat report *1990 National Assessment of Educational Progress Sampling and Weighting Procedures, Part 2: National Assessment, Final Report* (Rust, Burke, & Fahimi, 1992).

Chapter 6

PROCESSING OF MATERIALS AND DATA

Tillie Kennel, Lavonne Mohn, Linda Reynolds,
Dianne Smrdel, Brad Thayer, and Lynn Zaback

National Computer Systems

As a subcontractor to Educational Testing Service, National Computer Systems (NCS) was responsible for the printing, distribution, and receipt control of NAEP materials, the professional scoring of open-ended items, and the computer processing of assessment data. Detailed documentation of these activities is provided in the *Report of Processing and Professional Scoring Activities, NAEP - 1989-90* (Kennel, Mohn, Reynolds, Smrdel, Thayer, & Zaback, 1991).

This chapter summarizes the processing of 1990 NAEP materials and data, from the receipt of completed assessment materials at NCS to the transmittal of data to Educational Testing Service. A separate discussion of the professional scoring procedures and activities is given in Chapter 7.

Table 6-1 shows the numbers of 1990 NAEP assessment instruments that were received and processed at NCS for the 1990 national assessment and Trial State Assessment, and provides counts of documents that were scanned, key-entered, and professionally scored. A tremendous effort was required to process the large quantities of assessment materials. As shown in the table, more than 300,000 instruments were received for processing, and over three million open-ended item responses were read and marked by professional scorers.

The magnitude of the processing effort was augmented by the overlapping of the various processing activities as well as the concurrent completion of the winter portion of the main NAEP assessment and the 1990 Trial State Assessment. This required NCS to develop and implement flexible, innovative processing programs and a sophisticated process control system that allowed the integration of data entry and work flow management systems.

6.1 Processing Overview

Following a set of predetermined rules and specifications, NCS staff performed a variety of important procedures on materials received from the local administrators and assessment supervisors before releasing these materials into the NCS NAEP processing system. Control systems were used to monitor and control all NAEP materials returned from the field. The NAEP Process Control System contained the status of all sampled schools for all assessments

Table 6-1

1990 NAEP Assessment: Processing Totals

Sample	Number of Sessions	Student Booklets	Absentee Forms	Excluded Student Questionnaire	School Characteristics and Policies Questionnaire	Teacher Questionnaire	Scanned Documents	Scanned Sheets	Key-entered Documents	Open-ended Readings
Bridge										
Fall	835	17,382	1,850	1,077	268	0	9,597	145,870	10,980	109,682
Winter	865	16,386	1,331	1,116	266	0	8,731	108,565	10,368	110,752
Spring	1,138	22,939	5,013	1,239	262	0	19,089	278,727	10,364	154,145
Total	2,838	56,707	8,194	3,432	796	0	37,417	533,162	31,712	374,579
Main										
Age Class 1	1,363	32,509	2,471	2,332	491	1,033	35,157	255,418	3,679	296,866
Age Class 2	1,077	29,337	3,584	1,992	370	1,185	32,905	233,026	3,563	330,677
Age Class 3	1,077	28,376	6,534	1,446	267	0	33,212	247,544	3,411	347,919
Total	3,517	90,222	12,589	5,770	1,128	2,218	101,274	735,988	10,653	975,462
Trial State	3,950	100,879	6,457	5,738	3,538	10,949	124,023	2,345,006	3,538	1,840,100
Total	10,305	247,808	27,240	14,940	5,462	13,167	262,714	3,614,156	45,903	3,190,141

128

129

and their scheduled assessment dates. As materials were returned, the Process Control System was updated to indicate receipt dates, record counts of materials returned and document any problems discovered with the shipments. As sessions were processed, the system was updated to reflect the processed counts. Custom report programs were developed to allow ETS, Westat, and NCS staff to monitor the progress in the receipt control and processing operations.

A subsystem of the Process Control System, the Alerts System, was developed to record, monitor, and categorize all discrepant or problematic ("alert") situations. Throughout the processing cycle, alert situations were identified based upon the processing specifications. These situations were either flagged by computer programs or identified using clerical procedures. All situations that could not be directly resolved by the staff involved in the given process were documented through the NAEP Alert/Resolution Process. A form describing the problem was completed and an alert record was entered on the Process Control System. The information was then forwarded to project personnel for resolution. The Process Control System monitored the status of alerts from initial definition through assignment, resolution, and closing. Status reports that listed outstanding alerts were produced on demand. Upon resolution of the situation, the response was entered into the system with a code indicator. After the problem was corrected, the alert was closed by the Receiving Department.

NCS's Work Flow Management System was used to track individual sessions through every processing step, thus allowing project staff to monitor the status of all receipts and to locate materials for particular sessions, if necessary. The Work Flow Management System was also used by NCS to analyze the current work load, by project, across all work stations. By routinely monitoring this data, NCS's management staff were able to assign priorities to various components of the work and ensure that various phases of the data receipt and processing were in keeping with specified deadlines.

6.2 Document Receipt

6.2.1 Student Assessment Booklets

Shipments received from the supervisors went through several steps in the receiving area. The first was the receipt of the postcard that provided notification of a return shipment. The receipt of this card was entered into the Process Control System and the cards were filed.

When the shipment arrived, the receiving clerk checked the contents of the package and compared them to the accompanying transmittal. The receipt was entered into the Process Control System.

Receiving personnel removed the contents of the package, separated materials by school, and checked the contents for each school against the appropriate School Worksheets and Administration Schedules. If any discrepancies were discovered, an alert was issued and the session was held for further processing.

When a shipment had been thoroughly checked, the school number, session identifier, date and time of assessment, number of students to be assessed, number absent and number assessed by session were entered from the School Worksheet into the Process Control System.

Students absent from the assessment were recorded in the system using absent student forms, resembling booklet covers, which were batched together for scanning.

When a shipment had been verified as complete, the booklets were organized into work units and batched by session. The batch number was entered on the Work Flow Management System, facilitating the internal tracking of the session and allowing departmental resource planning. All student documents were forwarded to the professional scoring departme... The School Worksheets and Administration Schedules were forwarded to the NCS NAEP operations coordinator and filed for future reference.

6.2.2 Assessment Questionnaires

Two rosters were used to account for the distribution and return of all questionnaires: the Roster of Questionnaires for the School Characteristics and Policies and Excluded Student Questionnaires and the Roster of Teacher Questionnaires for the Teacher Questionnaires.

The excluded student questionnaires were compared to the Roster of Questionnaires and the Administration Schedule to verify demographic information, and were submitted for scanning as sufficient quantities became available for batching. The teacher questionnaires were checked against the Roster of Teacher Questionnaires and submitted to scanning when sufficient batching quantities were available. The school characteristics and policies questionnaire, a key entry document, was compared to the Roster of Questionnaires and the school number was verified. They were then batched and forwarded to the key entry department.

Two numbers were entered into the Process Control System for each type of questionnaire: number of questionnaires expected and number actually received. If any questionnaires were outstanding, the roster remained on file in the receipt area for check-in when the questionnaires arrived.

6.2.3 Booklet Accountability

In response to the sensitive issue of security, a booklet accountability system was used. Before distribution, all shipments that were sent to the supervisors were recorded on the Materials Distribution System. When the unused materials were returned at the end of the assessment, the number and type of documents were manually compared to the shipped quantities. Any major discrepancies were directed to Westat for follow-up. The unused materials were then inventoried and sent to storage.

6.2.4 Processing Reports

Eleven different receipt control status reports were produced by the NAEP Process Control System: Receiving Checkoff List, Receipt Control Status, Session Status, Questionnaire Status, Documents Processed by Form, Westat Receipt Control Errors, Receiving Receipt Control Errors, Scheduled Date Compare, Notification Date Compare, ETS Summary, and

Alert Status Report. These reports are described in detail in the *Report of Processing and Professional Scoring Activities.*

6.3 Data Transcription Systems

The transcription of the student response data into machine-readable form was achieved through the use of three separate systems: data entry, validation (pre-edit), and resolution.

6.3.1 Data Entry

Scannable Documents. The excluded student questionnaires, absent student forms, teacher questionnaires and the majority of the student booklets were scannable documents. The student booklets were batched by session, allowing each session to be tracked through the scoring and transcription process. The Process Control System monitored all sessions for a school.

The optical scanning devices and software used at NCS allowed a complete mix of NAEP scannable materials with no special grouping requirements. However, for manageability and tracking purposes, student documents, excluded student questionnaires, absent student forms, and teacher questionnaires were batched separately. In addition to scannable responses, the bar code identification numbers used to maintain process control were also decoded and transcribed to the NAEP computerized data file.

Each scannable NAEP document was uniquely identified by a number printed on each sheet of each document as it exited the scanner, permitting the data editors to quickly and accurately locate specific documents during the editing phase and providing a method for easy identification and retrieval of any document during the processing.

Data values from the booklet covers, reader identification fields, and item responses and scores were returned as numeric codes. Unmarked fields, fields with multiple marks, and fields from unreadable pages were specially coded for resolution staff to check and correct if possible.

When scanning was completed for each batch, the scanning program was terminated, which closed the dataset, terminated the link to the mainframe computer, and automatically submitted a computer job to run the pre-edit step. The scanned documents were then forwarded to the editing department for error resolution.

Key-entered Documents. A process of key entry and verification was used to capture the data from nine of the paced-tape booklets, the 18 reading and writing bridge booklets, the school characteristics and policies questionnaires, and the professional scoring reliability scoring sheets, which were all nonscannable documents. These data were entered using the Falcon online data entry system. Because the fields to be entered were titled to reflect the actual source document, all key entry fields were specific to the type of NAEP document being entered.

Document Definition. One of the more complex aspects of the NAEP project was the number of different documents that had to be processed. To do the proper edits, a detailed

document definition procedure was used to allow NCS to define an item once and use it in many blocks and to define a block once and use it in many documents. The procedure used was a document file that pointed to the appropriate blocks on a block file that point to appropriate items on an item file. Because the document was defined independently from the edit program, documents could be changed or added without changing the code in the edit program. The document, block, and item files are described in the *Report of Processing and Professional Scoring Activities.*

6.3.2 Data Validation

Each dataset produced by the scanning system contained data for one or more assessment sessions. These data had to be edited for type and range of response. The data entry resolution system was able to process materials from three age groups, three assessment types, one absent form, and five questionnaires simultaneously, as the materials were submitted to the system from scannable and nonscannable media.

The data records in the scan file were organized in the same order in which the paper materials were processed by the scanner. A record for each session header preceded all data records for that session; thus, the set of records belonging to one session was separated from the others by its session header record. The document code field on each record distinguished the header record from the data records.

When a batch header record was read, a pre-edit data file and an edit log were generated. As the program processed each record within a batch from the scan file, it wrote the edited and reformatted data records to the pre-edit data file and recorded all errors on the edit log. The fields on an edit log record identified each data problem by the batch sequence number, booklet serial number, section or block code, field name or item number, and data value. After each batch had been processed, the program generated a listing of the data problems and resolution guidelines, which was printed at the termination of the program. Details of the validation process are given in the *Report of Processing and Professional Scoring Activities*.

When the entire document had been processed, the completed string of data was written to the data file. When the next session header record was encountered, the program repeated the same set of processes for that session. When the program encountered the end of the file, it closed the dataset and generated an edit listing.

Accuracy checks were performed on each batch processed. Every 500th document of each booklet form was printed in its entirety, with a minimum of one document type per batch. This record was checked item by item against the source document for errors. No scanning errors were discovered through this process.

6.3.3 Data Editing

Throughout the system, high-quality procedures and software ensured that the NAEP data were correct. The initial editing that took place during the receipt control process included

verification of the schools and sessions. All student documents on the administration schedule were accounted for. The computer edits performed during data capture verified that each sheet of each document was present and that each field had an appropriate value. All batches entered into the system, whether key-entered or machine-scanned, were edited for errors.

Data editing, which took place after these checks, consisted of a computerized edit review of each respondent's document and the clerical edits necessary to make corrections based upon the computer edit. This data editing step was repeated until all data were correct.

The first phase of data editing was designed to validate the population and ensure that all documents were present. A computerized edit list, produced after NAEP documents were scanned or key-entered, and all the supporting documentation sent from the field were used to perform the edit function. The hard copy edit list included the number of students, school code, type of document, assessment code, error rates, suspect cases, and record serial numbers. Using these, the data editor verified that the batch had been assembled correctly, each school number was correct and all students documents within each session were present.

In the second phase of data editing, an experienced editing staff used a predetermined set of rules to review the field errors and record necessary corrections to the student data file. The same computerized edit list used in the first phase was used to perform this function. The editing staff reviewed the edit log prepared by the computer and the actual source document listed on the edit log as being "suspect" or containing possible errors. The corrections were identified by batch sequence number and field name for suspect record and field identification. The edit log indicated the current composition of the field. The editing staff checked this piece of information against the NAEP source document, flagging double grids, erasures, smudge marks, or omitted items.

If the error was correctable by the editing staff according to the editing specifications, the corrections were indicated on the edit listing. If an error was not correctable according the specifications, an alert was issued to the operations coordinator for resolution. When the correct information was obtained, the correction was indicated on the edit listing. If a suspected error was found to be correct as stated and no alteration was possible according to source documents and specifications, the programs were tailored to allow this information to be accepted into the data record and no corrective action was taken.

When the entire batch of sessions was resolved, the edit list was forwarded to the key entry staff, who entered and verified the corrections using the Falcon system. When all corrections were entered and verified for a batch, an extract program pulled the correction records into a mainframe dataset.

At this point, the post-edit program was begun. This program posted the corrections to specified records, and the edit criteria were again applied to all records. If there were further errors, another edit list was printed and the cycle began again. When the edit process had produced an error-free file, the booklet ID number was posted to the NAEP tracking file by age, assessment, school, and session, allowing the accurate accounting of the number of documents processed for a session within a school and the number of documents processed by form.

6.4 Transmittal of Data Files

The 1990 NAEP data collection resulted in seven classes of data files—student, school, teacher, absentee, excluded student, sampling weight, and item information. The structure and internal format of the 1990 NAEP database is a continuation of the integrated design originally developed by ETS in 1984. Data files containing the student and school information were sent to ETS and Westat at the conclusion of each assessment; all teacher information was transmitted at the end of the 1990 assessment period.

6.5 Document Storage

After the batches of documents had successfully passed the editing process, they were sent to the warehouse for storage. The storage locations of all documents were recorded on the inventory control system. Unused materials were sent to temporary storage until the assessment had been completed and the data files accepted, at which time the extra inventory was destroyed and a nominal supply of materials was stored permanently.

Chapter 7

PROFESSIONAL SCORING

Mary A. Foertsch, Claudia A. Gentile, Lynn B. Jenkins,
Chancey Jones, Lee R. Jones, and Beverly Whittington

Educational Testing Service

7.1 Introduction

The 1990 assessment included a variety of open-ended items—that is, items in which students were asked to produce a response, rather than select the correct answer from a series of options. Open-ended items were included in the main assessments in reading, science, and mathematics, the trial state assessment in eighth-grade mathematics, and the bridge assessments in reading, writing, and mathematics. The responses to these items were professionally scored by teams of trained readers.

Specifically, the 1990 main, bridge, and trial state assessments included the following numbers of professionally scored open-ended items:

Table 7-1
Numbers of Professionally Scored Open-ended Items, NAEP 1990

Subject Area	1990 Main and Trial State Assessments			
	Age 9/Grade 4	Age 13/Grade 8	Age 17/Grade 12	Trial State/Grade 8
Mathematics	41	42	48	35
Science	17	24	24	N/A
Reading	4	3	4	N/A

Subject Area	1990 Bridge Assessments		
	Age 9	Age 13	Age 17
Mathematics	44	27	57
Reading	5	8	9
Writing	6	6	6

The following sections summarize the scoring of open-ended items in the main, trial state, and bridge assessments. A more detailed discussion can be found in the *Report of Processing and Professional Scoring Activities, NAEP - 1989-90* (Kennel, Mohn, Reynolds, Smrdel, Thayer, & Zaback, 1991).

7.2 Main and Trial State Mathematics Assessments

7.2.1 Characteristics of the Scoring Guides

The mathematics portion of the 1990 national assessment included 97 discrete open-ended items designed to measure different aspects of students' mathematics understanding. Some of these items were administered at more than one grade level. In addition, a special paced-tape booklet at each grade contained estimation and innovative problem-solving items. The scoring guides for the 97 open-ended questions provided correct responses and solutions for each problem, alternative methods of solutions that could be used by students to solve the problems correctly, and certain incorrect student responses that indicated particular errors or misunderstandings.

The open-ended mathematics questions in the 1990 NAEP assessment measured five major content areas and three levels of mathematical ability that required students to demonstrate mathematical understanding and problem solving techniques at various levels of sophistication. The items required the scoring of computational answers, patterns, tables and charts, geometric figures, graphs, and brief explanation and justification statements. Each open-ended item had a unique scoring guide that provided the range of possible scores for the item and gave the criteria to be used in evaluating students' responses. To this end, each scoring guide was prepared and refined as an integral part of the item development process. At each step of the two-year NAEP item development and review process, the scoring guide for a given open-ended question underwent scrutiny and discussion and was revised as necessary. In addition, each scoring guide was subject to a final review and revision, if necessary, as part of the scoring process. While selecting sample responses for use in training readers, NAEP staff examined a variety of student responses to each question to determine the appropriateness of the scoring guide and to select illustrative student responses that were to be included in the packet of materials that was used to train readers.

In reporting results, student solutions to open-ended questions were reported as either correct or incorrect; no partial credit was reported. However, for the purposes of gathering information to assist in interpreting the results of the assessment, the following outline of score categories was used to score the items.

Outline for Scoring of Open-Ended Mathematics Items

Score	Definition
	Scores Awarded for Full Credit and Counted as Correct Responses:
8	This score indicated a correct answer in the detail required for commonly used methods of solution.
7	This score indicated a correct answer in the detail required for alternate solutions.
	Scores Awarded for Partial Solutions but Counted as Incorrect Responses:
5,6	These scores were given to responses that were correct to a point but were either incomplete or contained some error or irrelevant information.
	Scores Awarded No Credit and Counted as Incorrect Responses:
2,3,4	These scores indicated an incorrect response to the questions that clearly reflected a student's misunderstanding of the concept being measured or a commonly given incorrect response.
1	This score was given to responses that were either incorrect, indecipherable, irrelevant, or contained a statement to the effect that the student did not know how to do the task.
0	This score was given to questions for which there was no response.

7.2.2 Training

The readers were organized into five teams, each containing 12 readers and a team leader. In late January, 1990, before the training process began, the ETS test development staff and the team leaders prepared training sets (sets of sample responses to accompany the scoring guides) and refined the scoring guides. ETS mathematics specialists conducted the training of the readers with assistance from the five team leaders. Training involved explaining each item and its scoring guide and discussing responses that were representative of the various score points in the guide. When this was completed, the readers scored and discussed 5 to 20 randomly selected "practice papers" for each item, depending on the complexity of the item. The purpose of the training was to familiarize the readers with the scoring guides and to reach a high level of agreement among the readers. Following the group training, each reader on the teams scored all of the open-ended items in each of approximately 12 bundles of booklets, each of which contained an average of 27 booklets. During this practice, discussion sessions were held to review and clarify responses for which subjective judgment was required. In some cases the scoring guides were revised to track common incorrect answers that had not appeared in the sample sets. When the practice session was completed, the formal scoring process began. The initial training was completed in one week.

The items in the paced-tape booklets at all three grade levels were scored exclusively by the five team leaders. This enabled training for the large group of math readers to be shortened considerably from the two weeks that were scheduled to one week. Since the team leaders had been present during the process of reviewing the scoring guides and selecting the training samples, the time required to train the team leaders on the additional items was minimal.

7.2.3 Scoring

Materials from the 1990 Trial State Assessment and the winter sample of the 1990 national NAEP mathematics assessment were scored simultaneously and by the same readers. The scoring of the spring sample of the national assessment was done by two of the team leaders and two other readers.

To determine interrater reliability, 20 percent of the open-ended mathematics responses were scored by a second reader. The team leaders reviewed discrepancies between readers and initiated group or individual discussions when readers were having difficulty scoring certain items or when papers were given discrepant scores by different readers.

The percent score agreement between readers was high, above 95 percent, for a large majority of the questions. Table 7-2 shows the average percent agreement for all questions that were double-scored from the national winter and spring samples at all three grades, and for the Trial State Assessment at grade 8.

Table 7-2
Percent Agreement Between Readers, Mathematics Main and Trial State Assessments

	Winter-National		Trial State		Spring-National	
Grade	Mean Percent Agreement	Range of Agreement	Mean Percent Agreement	Range of Agreement	Mean Percent Agreement	Range of Agreement
4	96.8	84.0 - 99.5	—	—	98.3	92.9 - 100.0
8	97.1	87.8 - 100.0	96.9	86.8 - 99.4	97.9	90.8 - 99.7
12	96.2	90.6 - 99.5	—	—	97.5	90.0 - 99.7

7.3 Main Science and Reading Assessments

7.3.1 Characteristics of the Scoring Guides

Science. The 1990 main NAEP science assessment at grades 4, 8, and 12 contained 31 discrete open-ended items, in which students were asked to solve problems, interpret information or data, evaluate experimental procedures, and design experiments. Sixteen of the open-ended questions were figural response items in which students were asked to draw lines and arrows on figures and graphs in response to certain tasks. As with the mathematics assessment, the scoring guides were developed to provide diagnostic information. Questions

were scored on a scale of 0-8, with 0 representing no response and 8 representing a complete and correct response. Scores of 1-7 represented a variety of incorrect and partially correct responses. Because results were reported only as correct or incorrect, partially correct responses to certain questions were reported as correct.

Reading. The 1990 main NAEP reading assessment at grades 4, 8 and 12 contained seven discrete open-ended items. These items were designed to evaluate particular reading behaviors and aspects of comprehension. Each item was accompanied by a unique scoring guide that defined levels of success in accomplishing the task given.

7.3.2 Training

The training procedures used for the science and reading assessments closely resembled those described for the mathematics assessment. Working closely with NCS staff, ETS content area specialists managed the process of refining the scoring guides and selecting illustrative sample papers. Because the science and reading booklets were spiraled together, the training for the open-ended items in these subject areas was conducted consecutively.

7.3.3 Scoring

The procedures for scoring the science and reading items in the main assessment closely resembled those used to score the mathematics items. Twenty percent of the responses for each subject were double-scored to monitor interrater reliability. The percent score agreement between readers was high, above 90 percent, for a majority of the science questions and above 88 percent for a majority of the reading questions. Shown in the following tables are the average percent agreement between readers for all science questions (Table 7-3) and all reading questions (Table 7-4) that were double-scored in the national winter and spring samples at all three grades.

Table 7-3
Percent Agreement Between Readers, Main Science Assessment

	Winter		Spring	
Grade	Mean Percent Agreement	Range of Agreement	Mean Percent Agreement	Range of Agreement
4	92.4	77.9 - 97.6	91.8	74.6 - 98.9
8	91.6	78.7 - 99.7	91.2	77.4 - 99.7
12	90.2	70.3 - 99.5	91.0	76.1 - 100.0

Table 7-4
Percent Agreement Between Readers, Main Reading Assessment

Grade	Winter		Spring	
	Mean Percent Agreement	Range of Agreement	Mean Percent Agreement	Range of Agreement
4	88.4	77.7 - 93.6	88.2	83.6 - 93.9
8	89.0	87.5 - 93.4	90.4	88.3 - 92.7
12	87.3	83.8 - 88.9	88.4	85.8 - 91.3

7.4 Bridge (Trend) Assessments

7.4.1 Characteristics of the Scoring Guides

Mathematics. The open-ended items in the mathematics bridge assessment were scored on a right/wrong basis.

Science. Some open-ended science items were scored only to determine whether the student had responded to that item. Items that had not been reached were assigned a score of 0 and those in which the student had written something in the space provided were given a score of 1.

Reading. The scoring guides for the open-ended reading items focused on students' abilities to perform various tasks—for example, identifying the author's message or mood and substantiating their interpretation, making predictions based on given details, and comparing and contrasting information. The guides for the reading items varied somewhat, but typically included a range of scores, ranging from unsatisfactory to elaborated responses. Some of the reading items received secondary scoring based on the reactions or explanations the student provided.

Writing. All of the writing items for the three bridge assessments were scored using the primary trait method, which focuses on the writer's effectiveness in accomplishing the specific purpose of a given writing task. The primary trait scoring criteria defined five levels of task accomplishment: not rated, unsatisfactory, minimal, adequate, and elaborated. The scoring guide for each item described these levels in detail. In addition, a subset of the items were scored using holistic and mechanics methods, described later in this chapter.

7.4.2 Training

As with the main assessments, preparation for training readers of the bridge assessment materials entailed reviewing the scoring standards and selecting sample papers. However, several additional considerations were involved in the trend scoring. First, because it was necessary to train the 1990 readers to use the scoring standards from previous assessments (1984 for reading, 1986 for mathematics, 1988 for writing), close attention had to be given to

selecting sample papers that were congruent with the set used in the previous scoring. In addition, scores on booklets from the earlier assessments had to be masked so as to prevent readers from being influenced by the previous readers' scores.

During the training, the ETS subject area specialists and NCS table leaders reviewed the scoring guides, elaborated on the rationale underlying the scoring, and discussed illustrative sample responses. The readers then reviewed scored sample responses and conducted extensive practice scoring.

Because the mathematics items were scored as right, wrong, or omitted, lengthy training of readers was unnecessary. Readers were guided through the scoring rubrics, which listed the correct answer for the items in each of the blocks. Scoring began after a brief episode of practice scoring.

7.4.3 Scoring

After the initial part of the training but before scoring any 1990 bridge materials, 25 percent of the 1984 and 1988 materials were selected for practice scoring. When it was clear that a high degree of interrater agreement had been reached, the readers began scoring the 1990 trend assessment materials.

For the mathematics trend items, readers scored every open-ended response in each assessment booklet by gridding the appropriate score in the ovals provided at the bottom of the booklet page. Every tenth booklet was rescored to ensure that the items had been scored accurately. These quality control checks were monitored, and discrepancies were addressed and remediated. Most of the errors found during this check occurred as a result of scores not being gridded.

Two reliability studies were conducted for the reading and writing trend items. For the 1990 material, 25 percent of the open-ended responses were scored by a second reader to produce interrater reliability statistics. In addition, to ensure that readers were scoring in keeping with the 1984 and 1988 procedures, 20 percent of the 1984 reading booklets and 25 percent of the 1988 writing booklets were selected and intermixed with the 1990 booklets before the scoring began. Each booklet selected from the 1984 and 1988 assessments had the original score masked for each item scored. The readers marked their scores on separate sheets.

Interrater reliability in the reading and writing trend assessments was examined from two different perspectives.

First, to ensure that there was no "drift" in the interpretation of the scoring standards across time, it was important to study the extent to which the scores assigned by the 1990 readers were in agreement with scores assigned by readers in the previous scoring years. For reading, 25 percent of the 1984 papers were randomly selected and given to the 1990 readers to rescore. For writing, 25 percent of the 1988 papers were rescored. The results of these analyses indicated that the level of agreement tended to be fairiy high and there were no consistent shifts across time in the interpretation of the scoring criteria.

Second, it was important to know whether the 1990 readers understood and applied scoring standards consistently—in other words, whether they had a high level of agreement with one another. The results of this second set of analyses indicated that an acceptable level of interrater agreement had been achieved.

Tables 7-5 and 7-6 provide the interrater reliability figures for the reading and writing trend assessments.

Table 7-5
Percent Agreement Between Readers, Reading Trend Assessment

Age/Grade	1984 Papers Rescored in 1990		1990 Papers Scored Twice	
	Mean Percent Agreement	Range of Agreement	Mean Percent Agreement	Range of Agreement
9/4	92.6	89.6 - 95.7	81.1	70.9 - 87.3
13/8	79.0	70.7 - 85.2	70.1	64.5 - 77.4
17/11	93.5	90.7 - 97.0	78.6	71.3 - 83.5

Table 7-6
Percent Agreement Between Readers, Writing Trend Assessment

Age/Grade	1988 Papers Rescored in 1990		1990 Papers Scored Twice	
	Mean Percent Agreement	Range of Agreement	Mean Percent Agreement	Range of Agreement
9/4	85.0	76.9 - 91.1	82.5	75.2 - 92.1
13/8	82.9	75.4 - 92.6	76.4	66.1 - 86.8
17/11	78.1	71.6 - 85.5	77.8	71.8 - 84.6

7.5 Holistic Scoring

To gather information about changes across time in the fluency of students' writing, NAEP evaluated some of the 1990 trend writing responses using the holistic scoring method. In this approach, readers evaluate students' writing for general fluency, rather than focusing on specific aspects of writing achievement.

In the NAEP holistic scoring session, two writing trend items were scored from each of the three grade levels (at grade 4, "Flashlight" and "Spaceship"; and at grades 8 and 11, "Food on the Frontier" and "Recreation Opportunities"). The responses were drawn from three assessment years: 1984, 1988, and 1990.

7.5.1 Preparation

Before the holistic scoring began, the Chief and Assistant Chief Readers and the ETS writing specialist explained the theoretical underpinnings of holistic scoring to the table leaders and reviewed the structure of the 6-point scoring guides that were used. Only three scoring guides were needed, since the same guide was used for persuasive tasks at all three grade levels and likewise for informative tasks.

The group reviewed anchor papers chosen in 1988 to illustrate each point on the holistic scale. Then, for each item, the Chief Reader, Assistant Chief, and table leaders read 50 responses drawn from across the three assessment years and assigned scores to these papers. These selected papers were then collated and photocopied for use in training readers.

7.5.2 Training and Scoring

As the training began, the Chief Reader led the readers through a review of the prompt, the accompanying scoring guide, and the six anchor papers. The entire group and the individual tables then discussed the anchor papers at greater length. When the readers were clear as to the distinctions among the various score points, the Chief Reader chose a small set of responses from the training samples for the readers to score. Practice scoring (and subsequent discussion) continued until the Chief and Assistant Chief Reader were satisfied that the readers had reached a clear and highly uniform understanding of the scoring guide. When the scoring of one item was completed, the group began training for the next item.

7.6 Writing Mechanics Scoring

Mechanics scoring focuses on the extent to which the writer can control the conventions of written English (grammar, spelling, capitalization, punctuation).

7.6.1 Preparation

A subset of the responses from the 1990 assessment were selected for mechanics scoring. Every third booklet of the grade-eligible booklets was selected. Black students were oversampled to provide a more stable measure of their performance. In all, 1,772 responses were scored from the 1990 assessment. In addition, 10 percent of the papers that were previously scored for mechanics from the 1984 and 1988 assessments were rescored for reliability. This sample was selected by locating specific booklets from a list generated by ETS and resulted in a rescore of 300 papers for both years.

Before training the mechanics readers, the four table leaders met with the ETS subject area specialist. This group discussed the scoring guide, reviewed papers that were scored in 1988, and scored sample responses. After each participant independently scored a set of papers, the group reviewed the individual scores and compared them to the scores assigned in 1988. Discrepancies were discussed and resolved. The group then chose a subset of the scored papers to be included in a training packet for the entire group of readers.

7.6.2 Training

The training began with a detailed review of the scoring guide, which was organized into four sections: Type of Sentence Construction, Faulty Sentence Construction, Punctuation, and Word Level Categorization. Excerpts from reference sources were distributed and other guides mentioned. After discussing the guide, the group reviewed the training papers. When the group had practiced extensively, the actual scoring began.

7.6.3 Scoring

In scoring, the mechanics readers alternated among the different grade levels and assessment years. The mechanics readers marked each paper with a series of symbols, which designated each word or punctuation mark in error and indicated sentence type or faulty sentence construction. Each paper was scored independently by two different readers.

Resolution and quality control were conducted by a table leader who compared the scores marked on separate copies of the responses and resolved any discrepancies. Feedback was provided to each reader and follow-up training discussions were held as necessary. As the readers became more comfortable and proficient with the scoring, the table leaders became more extensively involved in the resolution and quality control processes. At that point, it was decided that it no longer was necessary for resolvers to read each paper and that comparing scores and resolving discrepancies would suffice.

Resolved packets were sent to the NCS word processing department where the text of the papers, along with the assigned scores and identification information, were entered into a computer according to carefully defined specifications. The scoring group proofread the data entry work against the scored papers and any discrepancies were resolved.

7.7 Portfolio Scoring

To gather information about the kind and quality of writing students regularly do in their English/language arts classes, NAEP conducted a special study of school-based writing, involving random samples of approximately 2,000 fourth-grade and 2,000 eighth-grade students. From the fourth graders, 1,066 papers were received (a 54 percent response rate) and from the eighth graders 1,059 papers were received (a 51 percent response rate).

Students papers were first descriptively analyzed and coded according to genre, length, audience, pre-writing, revision, use of sources, and teachers' comments. An evaluative analysis was then conducted on papers that fit into the three major genre categories: narrative, informative, and persuasive.

7.7.1 Preparation

Before training began on the portfolio submissions, the scoring coordinator solicited reviews from the portfolio committee and the scoring guide development team as well as from ETS assessment and subject area specialists. The scoring coordinator then met with ETS subject area specialists. This group discussed the scoring guide, reviewed the anchor papers, and aggregated comments from the other reviewers in order to revise the guide.

7.7.2 Training and Scoring

At the scoring session, a group of elementary and secondary teachers were trained on the informative scoring guide and then scored the informative papers. Next, they were trained on the narrative guide and then scored the narrative papers. Because only 50 papers in the persuasive category had been submitted, these were scored by the two chief readers.

As the training began, the training coordinator presented a definition of the genre in question, along with an explanation of the scoring guide. The chief reader at each of the two tables then led the scorers through a discussion of the guide and review of the anchor papers. When the scorers were clear as to the distinction among the various levels in the guide, the chief readers then gave the scorers a small set of responses from the training samples to score. Practice scoring and subsequent discussion continued until the chief readers and training coordinator were satisfied that the readers had reached a clear understanding of the scoring guide.

Papers were then scored, with the chief readers resolving any problems that arose. Twenty percent of the papers received a blind second scoring, so that interrater reliability could be estimated. The scores were then recorded on a tally sheet for entry into a computerized data analysis program.

Chapter 8

CREATION OF THE DATABASE, QUALITY CONTROL OF DATA ENTRY, AND THE 1990 NAEP DATABASE PRODUCTS

Alfred M. Rogers, David S. Freund, and John J. Ferris

Educational Testing Service

The data transcription processing of the 1990 NAEP data was conducted by the ETS subcontractor, National Computer Systems (NCS). The processes conducted by NCS, described in Chapter 6, resulted in the transfer to ETS of data files containing response data for students, teachers, and school administrators. At the same time, NCS transferred to ETS subcontractor Westat, Inc., the demographic data needed to derive sampling weights. Westat then provided ETS with data files that included sampling weights for assessed students, excluded students, and schools, principal questionnaire data, school characteristic data, and community characteristic data. Before any analysis could begin, these data files had to be merged into a comprehensive, integrated database. The creation of the database is described in section 8.1.

To evaluate the effectiveness of the quality control of the data entry process, the final database was sampled and verified in detail against the original instruments received from the field. The results of this procedure are given in section 8.2.

The integrated database was the source for the creation of three NAEP database products: the item information database, the restricted-use data files, and the secondary-use data files. These are described in section 8.3.

8.1 CREATION OF THE DATABASE

8.1.1 Merging Files

The transcription process resulted in the generation of nine sets of data files (bridge, first-half main, and second-half main for each of the three age/grade cohorts). Included in each set were student response data, school questionnaire data, and excluded student questionnaire data. A fourth data file contained teacher questionnaire data for the teachers of fourth- and eighth-grade students assessed in the first-half and second-half main assessments. The process of deriving sampling weights produced an additional three files (assessed students, excluded students, and schools) for each of the nine sets. Before data analyses could be performed, these files had to be integrated into a coherent and comprehensive database.

The database ultimately comprised three files per cohort: school, excluded student, and assessed student files. The student file contained data from all student samples--the main

assessment, the bridge to 1984, the bridges to 1986, and the long-term bridge. The student data were created in several steps, merging the student response data with the demographic and community variables, the students weights, and key derived variables such as parental education and home environment composites. Also, teacher data were appended to student records in cases where the student's teacher responded to a teacher questionnaire. In all steps, the booklet serial number was used as the matching criterion.

The school file was created by matching and merging the school questionnaire file with the school weights file and with a file of school variables supplied by Westat, which included demographic information about the school and the community. Some of these data were collected by Westat as part of the principal's questionnaire. The PSU and school code were used as the matching criteria. Each record of the resulting file was formed by merging the weight information with the response data and the school demographic and community data. Since not all schools returned their questionnaires, some of the output records contained only school identifying information and weight information. The school file could be accessed on its own or it could be linked to the assessed and excluded student files through the PSU and school codes.

The excluded student file was the result of merging the excluded student questionnaire file with the excluded student weights file. The booklet serial number was used as the matching criterion.

To ensure that the data were transferred accurately from NCS to ETS and that the processing control parameters used by analysis programs at ETS were properly in place, several quality control procedures were implemented. The principal procedure included matching independently generated frequency distributions computed at NCS with those that were produced at ETS, using ETS control parameters to process the data. Distributions for all student response questions (approximately 1,500 items) were verified to match.

When the appropriate files had been merged and proper quality control procedures had been completed, the database was ready for analysis. Any time that new data values, such as plausible values, were derived external to the database, they were added to the relevant files using the matching procedures described above. The secondary-use data files were later generated from this database.

8.1.2 Creating the Master Catalog

A critical part of any database is its processing control and descriptive information. A central repository of this information may be accessed by all analysis and reporting programs to provide correct parameters for processing the data fields as well as to provide consistent labeling to identify the results of the analyses. The NAEP master catalog file was designed and constructed to serve both of these purposes.

Each record of the master catalog contains the processing, labeling, classification, and location information for each data field in the NAEP database. The control parameters are used by the access routines in the analysis programs to define the manner in which the data values are to be transformed and processed.

Each data field has a 50-character label in the master catalog describing the contents of the field and, where applicable, the source of the field. The data fields with discrete or categorical values (e.g., multiple-choice items and professionally scored items, but not weight fields) have additional label fields in the catalog containing 8- and 20-character labels for those values.

The classification area of the catalog record contains distinct fields corresponding to predefined classification categories for the data fields. For a given classification field, a nonblank value indicates the code within that classification category for the data field. This permits the collection of identically classified items or data fields by performing a selection process on one or more classification fields in the catalog.

According to the NAEP design, it is possible for item data fields to appear in more than one student sample and in more than one block within each sample. The location fields of the catalog record contain the age, block and, where applicable, the sequence within the block for each appearance of the data field. (Fields such as plausible values and weights do not contain sequence numbers since these fields are not pertinent to a given block.)

The master catalog file was constructed in parallel with the collection and transcription of the assessment data to be ready for use by analysis programs when the database was created. As new data fields were derived and added to the database, their corresponding descriptive and control information were entered into the catalog.

One of the most important uses of the master catalog was the control of the creation of the secondary-use data files, codebooks, and file layouts. A synopsis of this process is presented in section 8.3.

8.2 QUALITY CONTROL OF DATA ENTRY

This was the first assessment for which data entry was done outside ETS, so the evaluation of the quality control of this process was of special interest. We are pleased to be able to report that the data entry was carried out with great success. As in past years, this NAEP database was found to be more than accurate enough to support the analyses that were done. Overall, the error rates found were comparable to those of past assessments.

This section includes evaluations of the data entry for both the national and trial state assessments. The reader is referred to *The Technical Report of NAEP's 1990 Trial State Assessment Program* (Koffler, 1991) for more details about the data entry for the state assessment.

The purpose of the analysis reported in this chapter was to assess the quality of the data resulting from the complete data entry system, from the actual instruments collected in the field to the final machine-readable database used in the analyses. The process involved the selection of instruments at random from among those returned from the field and the comparison of these instruments, character by character, with their representations in the final database. In this way, we were able to measure the error rates in the data as well as the success of the data entry system.

Of course, the observed error rate cannot be taken at face value. For example, the sample of school characteristics and policies questionnaires that happened to be selected at the national level for close inspection contained no errors at all. To conclude that the entire national school characteristics questionnaire database is therefore error free would be an act of extreme optimism; we may simply have been lucky with this particular random sample. What is needed is an indication of how bad the true error rate might be given what we observed. Such an indication is provided by confidence limits. Confidence limits indicate how likely it is that a value will fall outside a specified range of values in a specified context or distribution. In our analysis, the specified range is an error rate between zero and some maximum value beyond which we are confident that the true error rate does not lie; the specified context or distribution turns out to be the cumulative binomial probability distribution. An example should demonstrate this technique:

> Let us say that 1,000 booklets were processed, each with 100 characters of data transcribed for a total of 100,000 characters. Let us say further that five of these characters were discovered to be in error in a random sample of 50 booklets that were completely checked; in other words, five errors were found in a sample of 5,000 characters. While the observed error rate of .001 may be of some interest, our concern is rather with identifying an error rate that represents an upper limit on the true (unknown) error rate, with some degree of confidence. The following expression, deriving from the binomial theorem, may be used to establish the probability that the true error rate is .0025 or less:
>
> $$\sum_{j=0}^{5} \binom{5000}{j} \times .0025^{j} \times (1 - .0025)^{(5000-j)} = .0147$$
>
> This is the sum of the probability of finding five errors plus the probability of finding four errors plus. . . etc. . . plus the probability of finding zero errors in a sample of 5,000 with a true error rate of .0025—that is, the probability of finding five or fewer errors by chance when the true error rate is .0025. Notice that we did not use the size of the database in this expression. Actually, the assumption here is that our sample of 5,000 was drawn from a database that is infinite. The smaller the actual database is, the more confidence we can have in the observed error rate; had there been only 5,000 in the total database, our sample would have included all the data and the observed error rate would have been the true error rate. The result of the above computation allows us to say, conservatively, that .0025 is an upper limit on the true error rate with 98.53 percent (i.e., 1 - .0147) confidence; that is, we are quite sure that our true error rate is no larger than .0025.

The individual instruments are briefly discussed in the following sections and a summary table (Table 8-1) gives the upper 99.8 percent confidence limits for the error rates for each of the instruments as well as sampling rate information. The confidence limit of 99.8 percent was selected to make these results comparable to those of previous administrations when the same limit was used.

Table 8-1

Summary of Quality Control Error Analysis for 1990 NAEP Data Entry

Instrument	Entry Type	Different Booklets	Number of Booklets Sampled	Number of Characters Sampled	Number of Errors	Observed Error Rate	Upper 99.8% Confidence Limit
NATIONAL ASSESSMENT							
Student Booklets - Main	Scanned	63	195	25,098	20	.0008	.0015
Student Booklets - Bridges	Keyed	24	76	10,032	2	.0002	.0011
	Scanned	20	95	13,912	4	.0003	.0010
Excluded Student Questionnaire	Scanned	1	45	3,690	3	.0008	.0033
Teacher Questionnaire	Scanned	3	21	2,274	8	.0035	.0088
School Questionnaire	Keyed	1	39	8,786	0	zero	.0007
STATE ASSESSMENT							
Student Booklets	Scanned	7	280	40,600	11	.0003	.0006
Excluded Student Questionnaire	Scanned	1	29	2,262	7	.0031	.0082
Teacher Questionnaire	Scanned	1	103	8,652	22	.0025	.0045
School Questionnaire	Keyed	1	71	15,549	9	.0006	.0014

8.2.1 Student Data

About 146,000 students were assessed across all samples in the national assessment. Their booklets were sampled at the rate of one in 400, the same rate that was used in past assessments. Interestingly, the quality of the data that were key-entered appeared to be slightly better than the quality of the data from the optically scanned booklets. This difference was most apparent at age class 9 where no errors were found among the keyed data and 14 errors were found in the scanned books that were examined. The 9-year-old students seemed more likely to offer scanning "challenges," namely erasures and multiple responses that the scanner took to be legitimate answers. Even here, however, the error rate was not problematic for the analyses performed on the data.

The student data for the state assessment, some 100,000 books, were sampled at the slightly higher rate of one in 360 books in order to examine each of the seven different books in each of the 40 jurisdictions. The error rate was somewhat lower than the overall rate for the national assessment. However, it should be noted that the state assessment involved only 13-year-old students; at the national level, the error rate for 13-year-olds (not shown separately in the table, but the best of the three age classes) was comparable to that found in the state assessment. It appears that at this age, students were willing to work more carefully and hence were kinder to the optical scanner.

8.2.2 Excluded Student Questionnaire Data

A total of 9,178 questionnaires was scanned in the national assessment and 5,735 in the state assessment. These books were sampled at a rate of about one in 200 books. Although the same instrument was used in both the state and national assessments, the state data proved to be somewhat more challenging to the scanner. Again, it was for the most part a matter of erasures and multiple responses that the scanner took to be legitimate answers.

8.2.3 Teacher Questionnaire Data

There were 2,093 teacher questionnaires collected in the national assessment and 10,331 in the state assessment. One percent of these questionnaires were sampled for this quality control procedure. All the teacher questionnaires were scanned in this assessment, and the error rates were slightly higher than in past assessments in which they were keyed. Most of the errors involved misreadings of erasures and multiple responses by the scanner. However, a design flaw, which has since been corrected, caused some of the errors: In the science questionnaire, the classroom information was requested in a misleading format. As a result, a number of teachers responded in an area of the questionnaire not intended for such use and therefore not scanned for responses. The effect was to increase the incidence of missing data for three items:

T033601 - Are students assigned to this class by ability?
T033701 - What describes the ability of the students in this class?
T033801 - Which best describes the content of this course?

8.2.4 School Characteristics and Policies Questionnaire Data

There were 1,897 school characteristics and policies questionnaires collected in the national assessment and 3,536 in the state assessment. Two percent of these questionnaires were sampled. At the national level, no errors were found in the questionnaires that were checked; at the state level, nine errors were found. All of these errors involved confusions of numbers and percentages by the person responding to the questionnaire. In future assessments, questionnaires will not ask for percentages so that such confusion can be avoided.

8.3 NAEP DATABASE PRODUCTS

The NAEP database described to this point serves primarily to support analysis and reporting activities that are directly related to the NAEP contract. This database has a singular structure and access methodology that is integrated with the NAEP analysis and reporting programs. One of the directives of the NAEP contract is to provide secondary researchers with a nonproprietary version of the database that is portable to any computer system. In the event of transfer of NAEP to another client, the contract further requires ETS to provide a full copy of the internal database in a format that may be installed on a different computer system.

In fulfillment of these requirements, ETS provides three sets of database products: the item information database, the restricted-use data files, and the secondary-use data files. The contents, format and usage of these products are documented in the publications listed under the appropriate sections below.

8.3.1 The Item Information Database

The NAEP item information database contains all of the descriptive, processing, and usage information for every assessment item developed and used for NAEP since 1970. The primary unit of this database is the item. Each NAEP item is associated with different levels of information, including usage across years and age cohorts, subject area classifications, response category descriptors, and locations of response data on secondary-use data files.

The item information database is used for a variety of essential NAEP tasks: providing statistical information to aid in test construction, determining the usage of items across assessment years and ages for trend and cross-sectional analyses, labeling summary analyses and reports, and organizing items by subject area classifications for scaling analysis.

The creation, structure, and use of the NAEP item information database for all items used up to and including the 1990 assessment are fully documented in the NAEP publications, *A Guide to the NAEP Item Information Database* (Rogers, Barone, & Kline, 1990) and *A Primer for the NAEP Item Information Database* (Rogers, Kline, Barone, Mychajlowycz, & Forer, 1989).

The procedures used to create the 1990 version of the item information database are the same as those documented in the guide. The updated version of the guide also contains the subject area classification categories for the cognitive items.

8.3.2 The Restricted-use Data Files

The restricted-use data files are for the exclusive internal use of the NAEP contractor. They contain a complete copy of the internal NAEP respondent database in a structured, documented, and portable format.

The internal database is maintained in a compressed format to conserve computing resources and to increase analysis efficiency. The access methods developed for this database locate data fields dynamically during the execution of analysis programs. The restricted-use data files, on the other hand, are "rectangular" in structure; each data field is in the same location on every record within a file. This static data definition, while not efficient from a computing resource standpoint, is much easier to document and is not dependent on any computing machinery, operating system, or data access method.

The restricted-use data files serve several critical purposes. They provide an archive for all respondent data collected and derived for NAEP since 1970. They ensure compatibility of usage by expressing the data in consistent, rectangular formats. Their portability greatly facilitates transition of the respondent database to future NAEP contractors. The accompanying data file layouts and codebooks provide a standardized, comprehensive reference source for NAEP staff.

The contents and formats of the NAEP restricted-use data files are documented in the NAEP publication *A Guide to the NAEP Restricted-use Data Files* (Rogers, Barone, & Kline, 1989).

The procedures used to create the restricted-use data files for the 1990 assessment are the same as those used to create the secondary-use data files. Since the secondary-use data file distribution package contains more products, the generation procedures will be described in the following section.

8.3.3 The Secondary-use Data Files

The secondary-use data files are designed to enable any researcher with an interest in the NAEP database to perform secondary analysis on the same data as those used at ETS. They differ from the restricted-use data files in one important respect: all subregional identification information has been encrypted or excluded in order to maintain the confidentiality of the states, schools, and students who participated in the assessment.

The three elements of the distribution package are the data files, the printed documentation, and the microfiche copies of the assessment instruments. A set of files for each sample or instrument contains the response data file, a file of control statements that will generate an SPSS-X system file, a file of control statements that will generate a SAS system file, and a machine-readable catalog file containing control and descriptive information, intended for the user who does not use either SAS or SPSS-X. The printed documentation consists of four volumes: a guide to the use of the data files, and a set of data file layouts and codebooks for each of the three age cohorts (see *The NAEP 1990 Secondary-use Data Files User Guide* [Rogers, Kline, Johnson, Mislevy, Allen, & Rust, 1992]).

The remainder of this section summarizes the procedures used in generating the data files and related materials.

8.3.3.1 File Definition

There were two differences in the design of the 1990 assessment over that of the 1988 assessment that influenced changes to the definition of the secondary-use data files: the focused- BIB booklet design and the direct matching of teacher questionnaires to student assessment instruments. These changes, in turn, enhance the secondary user's ability to identify and analyze data of interest.

The focused-BIB design within the main assessment isolates the primary subject areas to separate groups of booklets. This permits the division of the main sample into subject-specific subsamples. The data files generated from these subsamples need only contain the data that is relevant to their corresponding subject areas and are therefore smaller and more manageable than their counterparts in previous assessments.

According to the design of the 1984, 1986, and 1988 assessments, only a sample of the teachers of the assessed students were asked to fill out the teacher questionnaires. The large size of the secondary-use main student files and the relatively low matching rate between students and teachers made it impractical if not physically prohibitive to produce a complete file with student and teacher information. Both the 1984 and 1986 secondary-use data packages had separate teacher data files which could be linked to the student data files for analysis. The teacher file in the 1988 secondary-use data package contained not only the teacher response data, but also the data from the students who could be matched to teacher questionnaires. This type of file was more appropriate for the analysis of teacher data as it defined the student as the unit of observation.

The intent of the 1990 assessment design was to collect data from the teachers of the main assessment students at certain age/grade levels who were administered science or mathematics booklets. A portion of the teacher questionnaire contained questions that were directly related to each matched student. This change in the design afforded a very high matching rate between student and teacher data. Therefore, for those subject areas in each age/grade cohort for which teacher data were collected, the teacher responses were appended to each student record in the secondary-use data files.

8.3.3.2 Definition of the Variables

The selection and arrangement of data fields, or variables, in each file was the next issue addressed. The initial step in this process was the generation of a LABELS file of descriptors of the variables for each data file to be created. Each record in a LABELS file contains, for a single data field, the variable name, a short description of the variable, and processing control information to be used by later steps in the data generation process. This file could be edited for deletion of variables, modification of control parameters, or reordering of the variables within the file. The LABELS file is an intermediate file only; it is not included on the released data files.

The first program in the processing stream, GENLYT, produced a printed layout for each file from the information in its corresponding LABELS file. These layouts were initially reviewed for the selection and ordering of the variables. Variables were excluded from secondary-use data files if they were classified as either confidential or nonapplicable.

The confidential variables included any descriptor or code that could be used to identify individual states, schools, or students in the NAEP sample. The PSU, school, teacher, and student identification codes used internally by ETS and Westat were "scrambled" according to specific algorithms to obtain new codes for use in linking the files together. These new codes were put on the secondary-use files in lieu of the original codes.

The nonapplicable variables were found mostly in the student database. In the database used for analysis and reporting, the bridge samples were combined with the main sample. Therefore, many of the variables that applied to the main sample students did not apply to the bridge sample students, and vice versa. Similarly, within the main assessment sample, students who were administered booklets in one subject area had no derived variables pertaining to the other subject areas. When the data for these samples were separated into different datasets for the secondary-use data files, these nonapplicable variables were excluded.

The variables on all data files were grouped and arranged in the following order: identification information, weights, derived variables, and response data. On the student data files, these fields were followed by the proficiency scale scores and teacher response data, where applicable. The identification information is taken from the front covers of the instruments. The weight data include sample descriptors, selection probabilities, and replicate weights for the estimation of sampling error. The derived data include sample descriptions from other sources and variables that are derived from the response data for use in analysis or reporting.

For each subject area of the main assessment, the item response data within each block were left in their order of presentation. The blocks, however, were arranged according to the following scheme: common background, subject-related background, and cognitive blocks in ascending numerical order. The responses to cognitive blocks that were not present in a given booklet were left blank, signifying a condition of "missing by design."

In order to process and analyze the spiral sample data effectively, the user must also be able to determine, from a given booklet record, which blocks of item response data were present and their relative order in the instrument. This problem was remedied by the creation of a set of control variables, one for each block, which indicated not only the presence or absence of the block but its order in the instrument. These control variables were included with the derived variables.

8.3.3.3 Data Definition

To enable the data files to be processed on any computer system using any procedural or programming language, it was desirable that the data be expressed in numeric format. This was possible, but not without the adoption of certain conventions for reexpressing the data values.

As mentioned in section 8.1, the responses to all multiple-choice items were transcribed and stored in the database using the letter codes printed in the instruments. This scheme afforded the advantage of saving storage space for items with 10 or more response options, but at the expense of translating these codes into their numeric equivalents for analysis purposes. The response data fields for most of these items would require a simple alphabetic-to-numeric conversion. However, the data fields for items with 10 or more response choices would require "expansion" before the conversion, since the numeric value would require two column positions. One of the processing control parameters on the LABELS file indicates whether or not the data field is to be expanded before conversion and output.

The ETS database contained special codes to indicate certain response conditions: "I don't know" responses, multiple responses, omitted responses, not-reached responses, and unresolvable responses, which included out-of-range responses and responses that were missing due to errors in printing or processing. The primary trait scores for the reading essay and writing items included additional special codes for ratings of "illegible," "off task," and nonrateable by the scorers. All of these codes had to be reexpressed in a consistent numeric format.

The following convention was adopted and used in the designation of these codes: The "I don't know" and nonrateable response codes were always converted to 7; the omitted response codes were converted to 8; the "not-reached" response codes were converted to 9; the multiple response codes were converted to 0; the "illegible" codes were converted to 5; and the "off task" codes were converted to 6. The out-of-range and missing responses were coded as blank fields, corresponding to the "missing by design" designation.

This coding scheme created conflicts for those multiple-choice items that had seven or more valid response options as well as the "I don't know" response and for those open-ended items whose primary trait scoring guide had five or more categories. These data fields were also expanded to accommodate the valid response values and the special codes. In these cases, the special codes were "extended" to fill the output data field: The "I don't know" and nonrateable codes were extended from 7 to 77, omitted response codes from 8 to 88, etc.

Each numeric variable on the secondary-use files was classified as either continuous or discrete. The continuous variables include the weights, proficiency values, identification codes, and item responses where counts or percentages were requested. The discrete variables include those items for which each numeric value corresponds to a response category. The designation of "discrete" also includes those derived variables to which numeric classification categories have been assigned. The open-ended items were treated as a special subset of the discrete variables and were assigned to a separate category to facilitate their identification in the documentation.

8.3.3.4 Data File Layouts

The data file layouts, as mentioned above, were the first user product to be generated in the secondary-use data files process. The generation program, GENLYT, used a LABELS file as input and produced a printable file. The LAYOUT file is little more than a formatted listing of the LABELS file.

Each line of the LAYOUT file contains the following information for a single data field: sequence number, field name, output column position, field width, number of decimal places, data type, value range, key or correct response value, and a short description of the field. The sequence number of each field is implied from its order on the LABELS file. The field name is an 8-character label for the field that is to be used consistently by all secondary-use data files materials to refer to that field on that file. The output column position is the relative location of the beginning of that field on each record for that file, using bytes or characters as the unit of measure. The field width indicates the number of columns used in representing the data values for a field. If the field contains continuous numeric data, the value under the number of decimal places entry indicates how many places to shift the decimal point before processing data values.

The data type category uses three codes to designate the nature of the data in the field: Continuous numeric data are coded "C"; discrete numeric data are coded "D"; open-ended item data are coded "O." Additionally, the discrete numeric fields that include "I don't know" response codes are coded "DI" and the open-ended items that include nonrateable response codes are coded "OI." If the field type is discrete numeric, the value range is listed as the minimum and maximum permitted values separated by a hyphen to indicate range. If the field is a response to a scorable item, the correct option value, or key, is printed. A range of correct options was indicated for those professionally scored items that were treated with cutoff scoring for IRT scaling. Finally, each variable was further identified by a 50-character descriptor.

8.3.3.5 Data File Catalogs

The LABELS file contains sufficient descriptive information for generating a brief layout of the data file. However, to generate a complete codebook document, substantially more information about the data is required. The CATALOG file provides most of this information.

The CATALOG file is created by the GENCAT program from the LABELS file and the 1990 master catalog file. Each record on the LABELS file generates a CATALOG record by first retrieving the master catalog record corresponding to the field name. The master catalog record contains usage, classification, and response code information, prefixed by the positional information from the LABELS file: field sequence number, output column position, and field width. Like the LABELS file, the CATALOG file is an intermediate file and is not included on the release data files.

The information for the response codes, also referred to as "foils," consists of the valid data values for the discrete numeric fields, and a 20-character description of each. The GENCAT program uses additional control information from the LABELS file to determine if extra foils should be generated and saved with each CATALOG record. The first flag controls generation of the "I don't know" or nonrateable foil; the second flag regulates omitted or not-reached foil generation; and the third flag denotes the possibility of multiple responses for that field and sets up an appropriate foil. All of these control parameters, including the expansion flag, may be altered in the LABELS file by use of a text editor, in order to control the generation of data or descriptive information for any given field.

The LABELS file supplies control information for many of the subsequent secondary-use data processing steps. The CATALOG file provides detailed information for those and other steps.

8.3.3.6 Data Codebooks

The data codebook is a printed document containing complete descriptive information for each data field. Most of this information originates from the CATALOG file; the remaining data came from the COUNTS file and the IRT parameters file.

Each data field receives at least one line of descriptive information in the codebook. If the data type is continuous numeric, no more information is given. If the variable is discrete numeric, the codebook lists the foil codes, foil labels, and frequencies of each value in the data file. Additionally, if the field represents an item used in IRT scaling, the codebook lists the parameters used by the scaling program.

Certain blocks of cognitive items in the 1990 assessment that are to be used again in later assessments for trend comparisons have been designated as nonreleased. In order to maintain their complete confidentiality, the descriptions of these items and their response categories have been substituted with generic labels in the file layouts, data codebooks, and user guide.

The frequency counts are not available on the catalog file, but must be generated from the data. The GENFREQ program creates the COUNTS file using the field name to locate the variable in the database, and the foil values to validate the range of data values for each field. This program also serves as a check on the completeness of the foils in the CATALOG file, as it flags any data values not represented by a foil value and label.

The IRT parameter file is linked to the CATALOG file through the field name. Printing of the IRT parameters is governed by a control flag in the classification section of the CATALOG record.

The LAYOUT and CODEBOOK files are written by their respective generation programs to print-image disk data files. Draft copies are printed and distributed for review before the production copy is generated. The production copy is printed on an IBM 3800 printer that uses laser-imaging technology to produce high-quality, reproducible documentation.

8.3.3.7 Control Statement Files for Statistical Packages

An additional requirement of the NAEP contract is to provide, for each secondary-use data file, a file of control statements each for the SAS and SPSS-X statistical systems that will convert the raw data file into the system data file for that package. Two separate programs, GENSAS and GENSPX, generate these control files using the CATALOG file as input.

Each of the control files contains separate sections for variable definition, variable labeling, missing value declaration, value labeling, and creation of scored variables from the

cognitive items. The variable definition section describes the locations of the fields, by name, in the file, and, if applicable, the number of decimal places or type of data. The variable label identifies each field with a 50-character description. The missing value section identifies values of those variables that are to be treated as missing and excluded from analyses. The value labels correspond to the foils in the CATALOG file. The code values and their descriptors are listed for each discrete numeric variable. The scoring section is provided to permit the user to generate item score variables in addition to the item response variables.

Each of the code generation programs combines three steps into one complex procedure. As each CATALOG file record is read, it is broken into several component records according to the information to be used in each of the resultant sections. These record fragments are tagged with the field sequence number and a section sequence code. They are then sorted by section code and sequence number. Finally, the reorganized information is output in a structured format dictated by the syntax of the processing language.

The generation of the system files accomplishes the testing of these control statement files. The system files are saved for use in special analyses by NAEP staff. These control statement files are included on the distributed data files to permit users with access to SAS and/or SPSS-X to create their own system files.

8.3.3.8 Machine-readable Catalog Files

For those NAEP data users who have neither SAS nor SPSS-X capabilities, yet require processing control information in a computer-readable format, the distribution files also contain machine-readable catalog files. Each machine-readable catalog record contains processing control information, IRT parameters, and foil codes and labels.

PART II

The Analysis of 1990 NAEP Data

Chapter 9

OVERVIEW OF PART II: THE ANALYSIS OF 1990 NAEP DATA[1]

Nancy L. Allen and Rebecca Zwick

Educational Testing Service

In 1990, NAEP conducted major assessments of reading, mathematics, and science. In each of these areas, the analyses included trend results providing links to previous assessments, as well as cross-sectional results for the 1990 assessment year, providing detailed information about student proficiency for grades 4, 8, and 12.

Another major component of the analyses of the 1990 data was the analysis of the writing assessment that links the 1990 trend results to results from previous assessments. For the first time, the 1990 writing assessment also included an analysis of writing portfolios. In addition, an assessment of mathematics for eighth-graders in 40 states and jurisdictions was analyzed. The details of the analysis of the data from the Trial State Assessment appear in a separate document, *The Technical Report of NAEP's 1990 Trial State Assessment Program* (Koffler, 1991).

Results from the analyses described in the following chapters were reported in *Trends in Academic Progress: Achievement of U.S. Students in Science, 1969-70 to 1990; Mathematics, 1973 to 1990; Reading, 1971 to 1990; and Writing, 1984 to 1990* (Mullis, Dossey, Foertsch, Jones, & Gentile, 1991); *Reading in School and out of School: Students' Literary Experience and Academic Achievement from 1988 to 1990 at Grades 4, 8, and 12* (Foertsch, 1992); *The State of Mathematics Achievement: NAEP's 1990 Assessment of the Nation and the Trial Assessment of the States* (Mullis, Dossey, Owen, & Phillips, 1991); *The 1990 Science Report Card: NAEP's Assessment of Fourth, Eighth, and Twelfth Graders* (Jones, Mullis, Raizen, Weiss, & Weston, 1992); and *The Writing Students Do in School: The 1990 NAEP Portfolio Study of Fourth and Eighth Graders' School-based Writing* (Gentile, 1992).

The data analysis for the 1990 NAEP assessments followed a number of steps in common across subject areas. These steps are described here, while later chapters explain some of the steps in detail and how the steps apply to analyses for the different subject areas.

---

[1]Eugene Johnson and John Mazzeo contributed to this chapter. Robert Mislevy, Norma Norris, and others provided helpful comments.

9.1 SAMPLES OF STUDENTS

The samples of students included in the 1990 NAEP assessment are listed and described in detail in Chapter 1. Only a brief description of the types of 1990 samples is given here. The 1990 samples were of three general types: bridge samples, the purpose of which was to provide links to earlier assessments; main NAEP samples, which were based on a common set of assessment procedures, including winter and spring administration times and calendar-year age definitions; and special study samples, which were from the same populations as the main NAEP samples but used different administration procedures. The winter administrations took place in January and February 1990; the spring administrations took place in March and April 1990.

The 1990 bridge assessments consisted of a bridge to 1984 in reading and writing, bridges to 1986 in reading, mathematics, and science, and a bridge for long-term trend in mathematics and science. *Trends in Academic Progress* contains, primarily, results from the analysis of the print bridge to 1984 (reading and writing) and the paced-audiotape bridge to 1986 (mathematics and science).

The 1990 main NAEP samples fell into two categories: focused-BIB and special studies. As described in Chapters 1 and 4, for each subject area, seven blocks of items were used to create seven different assessment booklets according to a focused balanced incomplete block (focused-BIB) design. The focused-BIB design provides for booklets that include three blocks of cognitive items in a single subject area, as well as background items. The blocks of cognitive items for reading, mathematics, and science included both multiple-choice and open-ended items; cognitive items in the writing blocks were all open-ended. Each block of cognitive items appeared in exactly three booklets. To balance possible block position main effects, each block appeared once as the first block of items, once as the second block, and once as the third block. In addition, the BIB design required that each block of items be paired in a booklet with every other block of items exactly once. Focused-BIB assessments were conducted for reading, mathematics, and science. Some additional booklets were included in the 1990 main assessment for special studies. The reading answer mode bridge for age 9/grade 4 falls into this category, as does the special assessment of mathematics estimation and higher-order thinking skills in an audiotape administration. The data for the reading answer mode bridge were collected in the same sessions as data for the focused-BIB assessments; however, the data for the assessment of estimation and higher-order thinking skills in mathematics were collected in separate sessions.

9.2 ANALYSIS STEPS

The analysis methods described in the following chapters are not necessarily identical across subject areas. The procedures used depend on whether test items are scored dichotomously (right vs. wrong) or ordinally and whether links across age groups or across assessments are required. Nevertheless, certain basic procedures are common to most or all of the analyses described in the following chapters; these are summarized here. In general, the procedures are listed in the order in which they occur for each subject-area analysis, although some overlap in the execution of the procedures sometimes takes place to reduce the total length of time to produce final results.

9.2.1 Preparation of Final Sampling Weights

Because NAEP uses a complex sampling design in which students in certain subpopulations have different probabilities of inclusion in the sample, the data collected from each student must be assigned a weight to be used in analysis. The 1990 NAEP weights were provided by Westat, Inc., the NAEP subcontractor in charge of sampling. Detailed information about the weighting procedures is available both in Chapter 10 and in a report prepared by Westat (Rust, Burke, & Fahimi, 1992).

9.2.2 Preliminary Analysis of Item Properties

Preliminary Item Analyses. The first analysis step for each data set, after the calculation of student weights, was to conduct item analyses within each age/grade cohort and within major reporting categories. These preliminary analyses had multiple purposes: to check the number of respondents, the scoring of items, and the coding of background data; to investigate the difficulty level of items and their ability to distinguish between students of high and low proficiency; to check for speededness; and to call attention to items that may have had popular but incorrect response options (indicating possible flaws in wording or scoring).

For each NAEP background item, the unweighted and weighted percent of students who gave each response were examined, along with the percent of students who omitted the item and the percent who did not reach the item. The number of respondents was also tabulated. Each block of dichotomously scored cognitive items was subjected to item analysis routines that yielded, for each item, the number of respondents, the percent of students who selected the correct response and each incorrect response, the percent who omitted the item, the percent who did not reach the item, and the correlation between the item score and the block score. In addition, summary statistics were computed for each block, including the alpha reliability (internal consistency). Although some blocks have few items, each block of items was analyzed separately, because the assessments are administered and timed block by block and because items from this assessment that are selected for use in future assessments will appear intact in the same block. Each student received three blocks of cognitive items.

The average proportion correct for the block is the average, over items, of the proportion of students who correctly answered each item. This is a measure of the difficulty level of the items. In all NAEP analyses (both conventional and IRT-based) of data collected in print administrations, a distinction is made between missing responses at the end of each block (i.e., missing responses subsequent to the last item to which a student provided an answer) and missing responses prior to the last observed response. In the former case, the item is treated as not having been presented to the student, or "not-reached." The latter type of missing response is treated as an intentional omission. In calculating the proportion correct for each item, students classified as not having been presented the item were excluded from the calculation of the statistic while students classified as intentionally omitting the item were treated as answering incorrectly. In NAEP analyses of data collected in paced-tape administrations, all missing responses are treated as intentional omissions.

The average biserial correlation is an indication of the ability of the items to distinguish between students of high and low proficiency. It is a measure of the average discrimination of

the items and is the average, over items, of the item-level biserial correlations (r-biserial). An item-level r-biserial is an estimate of the correlation between the response to the item (1 for correct, 0 for incorrect) and a criterion variable measuring proficiency. For each item-level r-biserial, the criterion variable for the correlation was the total block number-correct score, including the item in question. Students received zero points for all items that were not attempted, either at the end of each block or prior to the last observed response. Students classified as not receiving the block were omitted from the calculation of the statistic.

The proportion of students attempting the last item is one minus the proportion of students that received the block, but omitted the final item in the block. This proportion is often used as an index of the degree of speededness associated with the administration of a block of items.

The item analyses were later repeated after deleting items that were excluded from scaling. Summaries of the results for only the items that were included in the scales are presented in the following chapters.

Tables of Item-level Results. Tables of the percentages of students choosing each of the possible responses to each cognitive and background item within each of the samples administered in 1990 were created for use in writing the reports. The results for each item in the cross-sectional analyses were cross-tabulated against the basic reporting variables such as region, gender, race/ethnicity, and parents' education level. All percentages were computed using the sampling weights.

The sampling variability of all population estimates was obtained by the jackknife procedure used by ETS in previous assessments. In addition to having an estimator of the sampling variance of a statistic t, an indication of the number of degrees of freedom to attribute to the estimated variance is provided for selected key populations. The degrees of freedom of a variance estimator provide information on the stability of that variance estimator: The higher the number of degrees of freedom, the higher the stability of the estimator. Further details are given in Chapter 10.

Preliminary Analyses of Differential Item Functioning Across Age, Gender, and Racial/ethnic Categories. For subject areas that yielded dichotomous item responses, graphical techniques that are available through NAEP's modification of the BILOG computer program (Rogers & Nelson, 1990; Mislevy & Bock, 1982) were used to determine whether it was reasonable to assume a common item response function across age, gender, and racial/ethnic categories. In the present context, an item response function is the nonlinear regression of a dichotomous item response on an unobserved proficiency variable. In NAEP, this nonlinear regression is assumed to take the three-parameter logistic form described in Chapter 11. The NAEP-BILOG program produces plots that show the estimated item response function for a particular sample (say, the three age classes combined). In addition, NAEP-BILOG can plot expected proportions correct for specified subsamples (say, each of the three age classes) at several points along the proficiency scale (see Mislevy & Bock, 1989, for further discussion). The expected proportions correct can then be examined to determine whether departures from the common item response function are large or systematic. The same method can be used to check for differential item functioning (DIF) across gender and racial/ethnic groups. Items that functioned differently across groups were reviewed to determine whether they should be deleted.

In the case of items that function differently across age groups, another option that was used for some items was to estimate separate item response functions for each age level. The very small numbers of items for which DIF across age groups was identified in this preliminary stage are listed in subsequent chapters. No items were deleted due to DIF across gender or racial/ethnic groups.

Due to tight time constraints for producing subject-area reports, separate analyses of DIF using the Mantel-Haenszel procedure (Holland, 1985; Holland & Thayer, 1988) were conducted after results for the reports were available. The information from these analyses is available for item and test development for future assessments. The analyses are described in section 9.2.6, and summary results for each subject area are reported in subsequent chapters.

9.2.3 Scaling

Scales based on item response theory were derived for reading, mathematics, and science. In each subject area, a single scale was used for summarizing trends. One scale was used for the reading cross-sectional data; five and four subscales were created respectively for mathematics and science cross-sectional data. NAEP uses the methodology of multiple imputations ("plausible values") to estimate characteristics of the proficiency distributions. Chapter 11 describes in detail the theoretical underpinnings of NAEP's scaling methods and the required estimation procedures. Only the basic analysis steps are outlined here.

For developing scales in the dichotomously scored subject areas (all areas except writing), the steps were as follows:

1) Use the NAEP-BILOG program[2] (Mislevy & Bock, 1982) to estimate the parameters of the item response functions on an arbitrary scale, assuming the three-parameter logistic model.

2) Use the NAEP version of the MGROUP program (Rogers, 1991; Sheehan, 1985), which implements the method of Mislevy (see Chapter 11 or Mislevy, 1991) to estimate proficiency distributions for each student on an arbitrary scale, based on these item parameter estimates and the student's responses to cognitive items and background questions.

3) Determine the appropriate metric for reporting the results and transform the results as needed. This includes the definition of composite scores for mathematics and science where subscales were initially scaled, and the linking of current scales to scales from the past, as well as the selection of the mean and variance of new scales.

[2]NAEP-BILOG allows students in each of the three age classes and/or other groups to be designated as distinct populations. This is important because, in NAEP, item sampling is not random across age classes. In this situation, age class membership must be taken into account to obtain consistent item parameter estimates via marginal maximum likelihood (see Mislevy & Sheehan, 1989).

4) Use random draws from these proficiency distributions ("plausible values" in NAEP terminology) for computing the statistics of interest, such as means for demographic groups.

As explained in Chapter 11, the plausible values obtained through the IRT approach are not optimal estimates of individual proficiency; instead, they serve as intermediate values to be used in estimating subpopulation characteristics. Under the assumptions of the scaling models, these subpopulation estimates will be consistent, which would not be true of subpopulation estimates obtained by aggregating optimal estimates of individual proficiency.

The 1990 cross-sectional assessments of reading, mathematics, and science included a number of open-ended items. These items were included in analysis and reporting in two different ways. First, each item was examined individually. The percentage of students giving various responses to the items were produced and carefully inspected. Also, the items were dichotomized by subject-area experts so that right and wrong responses were available for each item. In this format, the items were included in the development of the proficiency scales. The following chapters provide information about the right/wrong scoring and the raters' reliability after categorizing responses into the two categories. Detailed information about the open-ended items can be found in Chapter 7.

As in reading, mathematics, and science, analyses of writing were conducted to determine the percentage of students who gave various responses to each writing cognitive and background question. However, unlike the other subject areas, a mean writing score procedure was used instead of the IRT procedure to produce overall estimates of writing achievement.

The mean writing score procedure calculates the mean score for each cognitive item in the subgroup of interest, and then calculates the mean across all cognitive items. This mean is then multiplied by 100 to convert it to a 0-to-400 metric. For example, as presented in *Trends in Academic Achievement*, the mean writing score for fourth-grade students in 1990 was 183. If one averages the six corresponding primary trait mean scores at grade 4 one obtains (2.12+1.80+1.88+1.59+1.86+1.76)/6, which is equal to 1.83. Multiplying by 100 results in the mean writing achievement of 183 at grade 4. This procedure is explained more fully in Chapter 15.

Scaling the Test Items. The data from both the trend and the cross-sectional samples were scaled. For all subject areas except writing, three-parameter logistic IRT scale score models were used. The long-term trend and the cross-sectional data were scaled separately. As described in section 9.2.2, in the course of the scaling, analyses of model fit to detect and correct aberrations related to differential functioning across subpopulations, or item-by-time interactions, were carried out.

Trend Scaling. To control for item parameter variation, item parameters for mathematics and science were reestimated using the data from 1990 as well as data from the 1986 and 1988 assessments. For reading, item parameters were reestimated using the data from 1990 as well as data from the 1984 and 1988 assessments. The resulting trend points based on these reestimated item parameters were then equated to the existing long-term trend lines by matching the mean and standard deviation of the 1986 proficiency distribution (1984 distribution, for reading), based on the reestimated item parameters, to the corresponding mean

and standard deviation based on the old item parameters. This process left the trend point estimates for the earlier years unchanged from their previously reported values and provided more stable estimates for the later trend points.

Additional trend samples were selected for the mathematics and science assessment in 1990 to supplement the samples that bridge to 1986. The items administered to these supplementary long-term trend samples were originally administered in early NAEP assessments, but had never before been administered in intact blocks or scaled. Investigation of the properties of these items revealed that the characteristics of the items differed from those of the items included in the bridge to 1986. The addition of the supplementary long-term trend items would have significantly changed the meaning of the proficiency scales, so the items administered to the supplementary long-term trend samples were not included in scaling. More details about the supplementary long-term trend items are included in Chapters 13 and 14.

Cross-sectional Scaling. Each of the cross-sectional assessments of mathematics, science, and reading have special characteristics that determine the procedures that were followed for cross-sectional scaling and that determine the relationship between the resultant scales and previously created scales for the subject area. The key consideration was the degree of similarity between the 1990 assessment of the subject area and earlier assessments in terms of the populations assessed and the characteristics of the assessment instrument used.

The cross-sectional assessment that was most different from previous assessments was the 1990 mathematics cross-sectional assessment. This assessment differed in a number of fundamental ways from the 1986 assessment, the assessment that formed the basis of the long-term trend scales for mathematics. The 1990 mathematics cross-sectional assessment differed from the 1986 assessment in sample age definition, time of testing, most of the items used, and most importantly, in the objectives that specify the content of the assessment. In comparison with the 1986 assessment, the objectives for the 1990 assessment de-emphasize the knowledge and skills process areas and the numbers and operations content area, increase emphasis on measuring problem-solving abilities, and incorporate the use of scientific calculators as an integral part of the assessment. Because of these changes, linking the 1990 mathematics scale to the 1986 scale would have been very misleading. Accordingly, new cross-sectional subscales were developed for mathematics.

Like mathematics, the 1990 cross-sectional assessment of science differs from the 1986 assessment (the basis of the science long-term trend scales) in sample age definition, time of testing and most of the items used. However, the objectives for the 1990 and 1986 science assessments are more similar to each other than are the 1990 and 1986 mathematics objectives. Because of the similarity in objectives and because the 1990 science assessment shares one block of items at each age/grade level with the 1986 science assessment, the scales based on the 1990 assessment were tied to the previous science scales through the shared items. It is important to note that while the objectives are similar for the two assessments, there are important differences. Also, it should be noted that the 1990 linking block of items at each age/grade level did not appear as an intact block of items in 1986, so the ability to tie the 1990 scales to the 1986 scales based on these items was affected by position and context effects between the two years. Furthermore, the small number of linking items taken by each student had a negative impact on the strength of the link. Chapter 14 describes the procedures used to link the two scales. It should be stressed that trend comparisons between 1986 and 1990 based on these

cross-sectional scale scores are largely meaningless due to the different age definitions and times of testing of the 1986 and 1990 assessed populations.

The 1990 assessment of reading shared the same age definitions, time of testing, mode of administration, and roughly half of the intact blocks of items with the 1988 cross-sectional assessment. Accordingly, a goal of the analysis of the cross-sectional reading data was to investigate linking the 1990 assessment to the 1988 cross-sectional scale, yielding a short-term trend from 1988 to 1990. There were two differences between the 1990 and the 1988 assessments that complicated the linking of the 1990 reading data to the 1988 data. First, NAEP has always excluded students for whom the assessment would have little meaning, such as students with severe physical or learning disabilities, or students with limited proficiency in English. The 1990 cross-sectional assessment incorporated new rules for the exclusion of students, which were meant to clarify and standardize the criteria for exclusion. If the change in the criteria for exclusion was found to have a substantial effect, comparisons between 1988 and 1990 would be of limited value, because the populations being compared would be different. The change in criteria for exclusion was found to have only minor impact, however. A second difference, for age 9/grade 4 students, was a change in the response mode format. In 1988, students gridded their responses directly in the test booklet. In 1990, a separate answer sheet was introduced at grade 4 (eighth- and twelfth-grade students used separate answer sheets in both 1988 and 1990). The change in answer mode was found to have a consistent effect. A separate sample of age 9/grade 4 students who responded in the test booklet was included in the 1990 cross-sectional assessment. This sample allowed the effect of this change to be examined, and made it possible to equate the two modes of response using an IRT-based common population equating procedure. Chapter 12 provides details of the steps taken to decide whether the linking of the 1990 reading results to the 1988 reading results was feasible, and to equate the two assessments.

For mathematics and science, the scaling was carried out within subscales. In determining the number and characteristics of the subscales to be created, it was necessary to balance two requirements:

1) the subscales should be as narrowly defined as possible to maximize the capability of identifying important interactions and to most nearly meet the assumptions of scaling, and

2) there must be a sufficient number of items within each subscale to support available scaling technology.

The subscale definitions were based on the content-by-process areas defined by the learning area committee and were guided by the above considerations. A verification of the relationships between the subscales (dimensionality analysis) was completed after scaling to inform interpretation of the subscales. A univariate scale was utilized for reading.

For mathematics and science, item parameter estimation was performed separately for each of the defined subscales, using data from all the age/grade samples for which the subscale was defined. Separate calibrations for subscales were not necessary for reading. Item parameter estimates on a provisional scale were obtained using the NAEP-BILOG program (Rogers & Nelson, 1990; Mislevy & Bock, 1982). The NAEP-BILOG item calibrations were

based on the data from a systematic sample of students in each sample (the calibration samples). The fit of the IRT model to the observed data was examined within each subscale by comparing the empirical item response functions (IRFs) with the theoretical functions. For each item presented at more than one grade, the empirical IRFs for each grade were compared. Any item for which the empirical IRFs differed significantly by grade was treated as a set of distinct items, a different item for each grade. Items that displayed significant lack of fit across one or more grades were omitted from scaling for those grades.

Generation of Plausible Values for Each Scale and Subscale. After the scales were developed, plausible values were drawn from the predictive distribution of proficiency values for each student. For the trend and reading scales, the plausible values were computed separately for each age or age/grade group and year and were based on the student's responses to the items going into the scale as well as on the values of a set of conditioning variables that were important for the reporting of proficiency scores. For the mathematics and science cross-sectional subscales, vectors of multivariate plausible values were drawn from the joint distribution of subscale proficiency values for the assessed student. These multivariate plausible values were computed separately for each age/grade and reflected the dependency between subscale proficiencies by utilizing shared variation among the subscales. All plausible values were later rescaled to the final proficiency scale metric using appropriate linear transformations.

The variables used in conditioning a given cross-sectional scale or group of subscales included a broad spectrum of background, attitude, and experiential variables and composites of such variables and explicitly included all standard reporting variables. The set of conditioning variables were defined with the aim of holding to low levels secondary biases in analyses involving a broad range of variables not included in the conditioning model. To minimize potential convergence problems for the mathematics and science cross-sectional scales, the original background variable contrasts were standardized and transformed into a set of linearly independent variables by extracting principal components from the correlation matrix of the original contrast variables. The principal components, rather than the original variables, were used as independent variables in the conditioning model for those scales. Trend scales used the same or similar sets of conditioning variables that were used when the scales were originally constructed. Details of the conditioning process and of the NAEP-MGROUP computer program that implements the process are presented in Chapter 11.

Definition of Composites for the Multivariate Scales. In addition to the subscale plausible values, composites of the subscales for the mathematics and science cross-sectional assessments were created as a measure of overall proficiency within each subject area. These composites were weighted averages of the subscale plausible values, where the weights reflected the relative importance of the subscales provided in the specifications of the learning area committee, and were approximately proportional to the number of items in each subscale at a given age.

Transformation to the Reporting Metric. For each trend or cross-sectional scale or subscale, a set of linear transformation coefficients were obtained to link the scale to previous scales or to an arbitrary scale, usually having a mean of 250.5 and a standard deviation of 50. Mathematics and science trend scales were equated to the 1986 cross-sectional scales. The

reading trend scale was equated to the 1984 cross-sectional scale. The transformations were of the form

$$\theta_{proficiency} = \alpha + \beta\theta_{calibrated}$$

where

$\theta_{proficiency}$	=	scale level in terms of the system of units of the final proficiency scale used for reporting
$\theta_{calibrated}$	=	scale level in terms of the system of units of the provisional NAEP-BILOG scale
β	=	$(SD_{new} / SD_{calibrated})$
α	=	$(M_{new} - \beta M_{calibrated})$
SD_{new}	=	the estimated or selected standard deviation of the proficiency distribution to be matched
$SD_{calibrated}$	=	the estimated standard deviation of the sample proficiency distribution on the provisional NAEP-BILOG scale
M_{new}	=	the estimated or selected mean of the proficiency distribution to be matched
$M_{calibrated}$	=	the estimated mean of the sample proficiency distribution on the provisional NAEP-BILOG scale

The final transformation coefficients for transforming each provisional scale and subscale to the final reporting scale are given in subsequent chapters.

Tables of Proficiency Means and Other Reported Statistics. Proficiencies and trends in proficiencies were reported by age for a variety of reporting categories. Additionally, the percentages of the students within each of the reporting groups who are at or above the scale anchor points were reported to provide information about trends in the distribution of achievement within each subject area. The variances of all estimates based on proficiency values included the component due to the error due to the latency of proficiency values of individual students as well as the component due to sampling variability.

9.2.4 Drawing Inferences from the Results

Similar to the manner in which the standard error for an individual group mean or proportion is used, the standard error of the difference between means or proportions can be

used to determine whether differences between assessment years or subgroups are statistically significant. If one wants to hold the certainty level for a specific set of comparisons at a particular level (e.g., .95), then multiple-comparisons procedures need to be used. One such procedure, the Bonferroni method, was used to form conservative confidence intervals for the trend differences between each previous assessment year and 1990.

Additionally, tests for linear and quadratic trends were applied to some of the national trend data in reading, mathematics, and science. The linear and quadratic components of the trend in average proficiency for a given subject area and age group were estimated by applying two sets of orthogonal contrasts to the set of average proficiencies by year. The linear component of the trend was estimated by the sum $b_1 = \Sigma c_j X_j$, where the X_j are the proficiency means by year and the c_j are defined so that b_1 corresponds to the slope of an unweighted regression of the proficiency means on the assessment year. The quadratic component was estimated by the sum $b_2 = \Sigma d_j X_j$, where the d_j are orthogonal to the c_j and are defined such that b_2 is the quadratic term in the unweighted regression of the proficiency means on the assessment year and the square of the assessment year. The statistical significance of b_1 and b_2 was evaluated by comparing each estimate to its standard error. The standard error of b_1 was computed as the square root of the sum $\Sigma c_j^2 SE_j^2$, where SE_j is the standard error of X_j. The standard error of b_2 was analogously defined. Tests for linear and quadratic trends allow conclusions about the patterns in the means for several points in time or for several related subgroups of students.

9.2.5 Teacher Questionnaires

Teachers in the sampled schools who taught students who were in the age 9/grade 4 mathematics main BIB sample or who taught students who were in the age 13/grade 8 mathematics or science main samples were asked to complete a two-part questionnaire about their own teaching background and the classroom practices they employ. The first part of the questionnaire pertained to the teacher's background and training. The second part pertained to the procedures used by the teacher for *each class* containing an assessed student. (See Chapter 2 for a description of the teacher questionnaires.)

To analyze the data from the teacher questionnaires with respect to the students' data, each teacher's questionnaire had to be matched to all of the sampled students who were taught mathematics or science by that teacher. In the subsequent chapters two separate match rates are given. The first is the percentage of students that could be matched to both the first and second parts of the teacher questionnaire. For these students, information is available not only about the background and training of their mathematics or science teachers, but also about the methods used in the particular mathematics or science class they attended. The second match rate is the percentage of students that could be matched to the first part of the teacher questionnaire. This match rate is larger because more students could be matched with information about a teacher than with information about the particular class they attended. Note that these match rates only reflect the student-level missing data. They do not reflect the additional missing data due to item-level nonresponse on the part of teachers. Variables derived from the teacher questionnaires were used as reporting variables at the student level and as variables that contributed to conditioning for the appropriate samples.

9.2.6 Scale Anchoring

The meaning of proficiency scales can be described either in a norm-referenced fashion, by providing information about how well students do in relationship to one another, or in a criterion-referenced fashion, by providing information about the correspondence between what a student does and the continuum of proficiency scores. Scale anchoring is a process that NAEP has used, beginning with the 1984 reading scale, to improve the utility of proficiency scale results by providing a criterion-referenced interpretation of selected scale levels. In this way, NAEP can further its goals of describing what students know and can do and stimulating debate about whether these levels of performance are satisfactory.

The report for reading, mathematics, and science trend results includes scale anchoring information as an aid to their interpretation. The anchoring process identifies items, if any, that a vast majority of students at a selected scale level can answer correctly but that most students at lower levels cannot. Such items are then reviewed by subject area specialists to interpret what differentiates among the score levels. The result is descriptions of student proficiency at each of the levels and a set of selected items that exemplify the interpretation. (See Beaton, 1987, and Johnson, 1988, for details of the anchoring process.) In addition to this information, subject-area reports include all of the released items (those items not held secure for use in future assessments) that anchored at a specified level.

Reading, mathematics, and science had been anchored previously. Since the trend scales are based on items previously scaled and anchored, the anchor-point descriptions previously developed hold for the 1990 trend scales. These descriptions also hold for the reading and science cross-sectional scales, because they were tied to 1986 scales.

The 1990 mathematics cross-sectional composite was anchored using techniques similar to those used for the 1986 mathematics assessment. The composite, rather than subscales, was anchored because there was an insufficient number of items within each subscale to support the scale anchoring process. In NAEP's scale anchoring process, the first step is to choose four to five scale points to be anchored. For each point, items are then evaluated as potential anchor items, based on the percent of correct responses among students with proficiency levels at that point, as well as the corresponding percent for the next lower anchor point. For the anchoring of the 1990 mathematics composite scale, an item was considered to anchor at a particular point if (1) the percent of students with proficiency levels approximately equal to that point (i.e., within a 25-point interval centered at the point) who answered correctly was at least 65, (2) the percent of students with proficiency levels approximately equal to the *next lower anchor point* who answered correctly was less than 50, (3) the difference between the percents in (1) and (2) was at least 30, and (4) the estimated percent correct at the anchor point and the next lowest anchor point was based on at least 100 students. (Of course, conditions (2) and (3) did not apply to the lowest anchor point.) After the items that anchored at each point were determined, subject area experts chose from among these the items that best characterized each point and developed descriptions of the anchored proficiency levels. The descriptions provide information about the types of skills that are possessed by a large proportion of students at that anchor point, but are not possessed by most students at lower levels. The percents of students at or above each anchor level are given in NAEP subject-area reports, along with the exemplar items and scale-point descriptions. As a check on the generalizability of the interpretation process, the anchored proficiency levels were independently described by two groups of mathematics

experts, each consisting of 10 members. Upon comparing the results, both groups agreed that the two sets of interpretations were very similar and that either set would have appropriately described the anchor item information. The two groups then worked together to obtain the final interpretations. Additional details about the anchoring of the 1990 mathematics cross-sectional composite are included in Chapter 13.

9.2.7 Differential Item Functioning (DIF) Across Demographic Groups

To assess DIF for two groups, examinees from the focal group (usually a minority group or females) were matched to examinees from the reference group (usually a majority group or males) who have the same overall test score. Performance on each item was then compared within these score categories. This is true in the examination of item response functions used in the preliminary examination of items and in the approach to DIF analysis that was taken after the scaling process was completed, based on the Mantel-Haenszel test (Holland, 1985; Holland & Thayer, 1988). Identifying items as differentially functioning for any particular demographic groups using either of these methods gives no indication of the fairness of the items.

Application of the Mantel-Haenszel approach to NAEP data is less straightforward than in conventional testing programs where all examinees receive the same set of items. In NAEP, we can apply the Mantel-Haenszel procedure either at the block level or at the booklet level, using scores on the block or booklet as the matching variable. At the block level, sample sizes would be larger, whereas at the booklet level, the matching criterion would be more reliable because it would be based on a larger number of items. Subsequent to investigating the advantages and disadvantages of each of these approaches, the Mantel-Haenszel method was applied to blocks of data from the reading, mathematics, and science cross-sectional assessments to investigate DIF across several demographic groups of interest. When the sample sizes could support them, comparisons were made of male and female students, White and Black students, and White and Hispanic students.

The Mantel-Haenszel procedure is performed separately for each item and the value of a test statistic, the Mantel-Haenszel D-DIF statistic, is computed. In the application of the ETS corporate-wide guidelines for interpretation of the results of the Mantel-Haenszel procedure, the size and statistical significance of the Mantel-Haenszel D-DIF statistics are used to classify items as "A" items, which are considered free of DIF, "B" items, which show some evidence of DIF, or "C" items, which show notable evidence of DIF (Zwick & Grima, 1991; Zwick & Ercikan, 1989). ETS guidelines require an examination of "C" items by content and sensitivity experts to make judgments about the fairness of these items. In the 1990 NAEP analyses, "C" items were examined by a committee of subject-area and testing experts to determine whether they should be retained for use in future assessments.

Procedures for DIF analysis are less clear-cut in the case of the writing assessment, in which students received from one to three ordinally scored items. (Similar considerations apply to the open-ended items in other subject areas. However, for the 1990 analysis these items were dichotomized.) The development of chi-square procedures for DIF investigations on these items is being researched.

9.2.8 Dimensionality Analysis

After scaling the mathematics and science cross-sectional data, the dimensionality of the item pools was investigated. The results of the dimensionality analyses provide information for test development and the development of future scales, as well as for interpreting the 1990 scales.

Dimensionality was investigated using information about the covariance structure to estimate the disattenuated correlations between content area subscales. These disattenuated content area intercorrelations were estimated from interfactor correlations from a confirmatory factor analysis where each content area defined a separate factor. That is, two testlets were assembled within each content area by splitting each content area subscale into odd-even halves within each booklet. The odd and even sets of items within a subscale were simply scored as number correct. Then a model with the same number of factors as the number of subscales was identified by its appropriate pair of odd/even scores, so that the model was based on the content area subscales. More specifically, if there had been three subscales, the factor pattern shown in Table 9-1 would have been estimated using the "asymptotically distribution free best general least squares estimators" in LISREL (Joreskog & Sorbom, 1986). The asterisks in the table refer to factor loadings to be estimated while the zeros refer to loadings that are constrained to be zero.

Table 9-1
Factor Pattern for a Three-factor (Subscale) Model

		Factors		
Subscale	**Item Set**	**1**	**2**	**3**
1	Odd	*	0	0
1	Even	*	0	0
2	Odd	0	*	0
2	Even	0	*	0
3	Odd	0	0	*
3	Even	0	0	*

Given the above constraints, the intercorrelation matrix among the factors provides estimates of the correlations among subscales corrected for reliability. This model was replicated across the seven main BIB test booklets for each age/grade level and each subject area. The subscale intercorrelations averaged across the seven booklets are presented in Chapters 13 and 14 for mathematics and science, respectively.

9.3 OVERVIEW OF CHAPTERS 10 THROUGH 15

The remaining chapters in Part II of this report are as follows:

Chapter 10: The 1990 National Assessment used a stratified multistage probability sampling design that provided for sampling certain subpopulations at higher rates (see Chapter 3). Because probabilities of selection are not the same for all assessed students, sampling weights must be used in the analysis of NAEP data. Also, in NAEP's complex sample, observations are not independent. As a result, conventional formulas for estimating the sampling variance of statistics are inappropriate. Chapter 10 describes the weighting procedures and methods for estimating sampling variance that are necessitated by NAEP's sample design. Further detail on sampling and weighting procedures is provided in *1990 National Assessment of Educational Progress Sampling and Weighting Procedures, Part 2: National Assessment, Final Report* (Rust, Burke, & Fahimi, 1992), a report prepared by Westat, Inc., the NAEP subcontractor in charge of sampling.

Chapter 11: A major NAEP innovation introduced by ETS is the reporting of subject-area results in terms of proficiency scales. Scaling methods can be used to summarize results even when students answer different subsets of items. For purposes of summarizing dichotomous item responses, NAEP developed a scaling technique that has its roots in item response theory and in the theories of imputation of missing data. Chapter 11 describes this scaling technique, the underlying theory, and the application of these methods to 1990 NAEP data. The final section of Chapter 11 gives an overview of the NAEP scales that were developed for the 1990 assessment.

Chapter 12: The two main components of the 1990 reading analysis are described in this chapter. First, the reading trend results for the years 1971 through 1988 were extended to include 1990 at ages 9, 13, and 17. The results of the reading trend analysis, which include the percents of students at or above the reading scale anchor points established in 1984, are reported in *Trends in Academic Progress: Achievement of U.S. Students in Science, 1969-70 to 1990; Mathematics, 1973 to 1990; Reading, 1971 to 1990; and Writing, 1984 to 1990* (Mullis, Dossey, Foertsch, Jones, & Gentile, 1991). In addition, a detailed cross-sectional analysis of reading was conducted for grades 4, 8, and 12 in 1988, including a study of the association between reading proficiency and student background variables. The cross-sectional analyses are reported in *Reading in School and out of School: Students' Literary Experience and Academic Achievement from 1988 to 1990 at Grades 4, 8, and 12* (Foertsch, 1992). The effect of changing from answering in the test booklet to answering on a separate answer sheet was studied for age 9/grade 4 students.

Chapter 13: The trend and cross-sectional analyses of the mathematics data are detailed in Chapter 13. The results of the trend analysis, which provided links from 1976 through 1990 for ages 9, 13, and 17, are reported in *Trends in Academic Progress*. A detailed cross-sectional analysis for grades 4, 8, and 12 in 1990 was also conducted, including an examination of the association of mathematics knowledge with instructional techniques and student background variables. The cross-sectional results, which also include the percents of students at or above anchor points determined in 1990, appear in *The State of Mathematics Achievement: NAEP's 1990 Assessment of the Nation and the Trial Assessment of the States* (Mullis, Dossey, Owen, & Phillips, 1991). At grades 4 and 8, background information and data on instructional methods were

collected from teachers and the relation of these variables to mathematics proficiency was examined.

Chapter 14: Like the reading and mathematics analyses, the science analysis consisted of two main components. The science trend results, which provide a link to 1970, 1973, 1977, 1982, and 1986, are reported in *Trends in Academic Progress*. A detailed cross-sectional analysis of science for grades 4, 8, and 12 in 1990 was also conducted, including an examination of the association of science skills with instructional techniques and student background variables. For grade 8, teacher data were collected and their association with science proficiency was analyzed. The cross-sectional results are reported in *The 1990 Science Report Card: NAEP's Assessment of Fourth, Eighth, and Twelfth Graders* (Jones, Mullis, Raizen, Weiss, & Weston, 1992).

Chapter 15: The writing analysis consisted only of a bridge assessment. The writing trend results, which provide a link to 1984 and 1988 for grades 4, 8, and 11, are reported in *Trends in Academic Progress*. A special part of the bridge assessment was a sample of writing portfolios gathered from the students in the assessment samples. Results for the portfolio analysis are presented in *The Writing Students Do in School: The 1990 NAEP Portfolio Study of Fourth and Eighth Graders' School-based Writing* (Gentile, 1992).

Chapter 10

WEIGHTING PROCEDURES AND ESTIMATION OF SAMPLING VARIANCE[1]

Eugene G. Johnson

Educational Testing Service

Keith F. Rust

Westat, Inc.

Neal Thomas

Educational Testing Service

As was the case in previous assessments, the 1990 National Assessment used a complex sample design with the goal of securing a sample from which estimates of population and subpopulation characteristics could be obtained with reasonably high precision (as measured by low sampling variability). At the same time, it was necessary that the sample be economically and operationally feasible to obtain. The resulting sample had certain properties that had to be taken into account to ensure valid analyses of the data from the assessment.

The 1990 NAEP sample was obtained through a stratified multistage probability sampling design that included provisions for sampling certain subpopulations at higher rates (see Chapter 3). To account for the differential probabilities of selection, and to allow for adjustments for nonresponse, each student was assigned a sampling weight. Section 10.1 discusses the procedures used to derive these sampling weights.

Another consequence of the NAEP sample design is its effect on the estimation of sampling variability. Because of the effects of cluster selection (students within schools, schools within primary sampling units) and because of the effects of certain adjustments to the sampling weights (nonresponse adjustment and poststratification), observations made on different students cannot be assumed to be independent of one another. As a result, ordinary formulas for the estimation of the variance of sample statistics, based on assumptions of independence, will tend to underestimate the true sampling variability. Section 10.2 discusses the jackknifing

[1]The statistical programming for this chapter was provided by Yim Fai Fong, Bruce Kaplan, Phillip Leung, and Wing Lowe of Educational Testing Service, and Dalia Kahane and Lana Ryaboy of Westat, Inc.

technique used by NAEP to estimate sampling variability. (The estimation of variability due to imperfect measurement of individual proficiency is discussed in Chapter 11.)

The jackknifing technique provides good quality estimates of sampling variability but requires considerable computations. Section 10.3 suggests the use of design effects, combined with conventional variance estimation formulas, as a simple approximation to sampling variability.

Yet another effect of the multistage NAEP sampling scheme is an increase in the variability of variance estimates, as compared with directly drawing independent samples of students (or of schools) without clustering them. Assuming that the distribution of variance estimates is approximately chi-square, the variability can be taken into account by specifying the degrees of freedom of the approximating chi-square variable: the higher the degrees of freedom, the lower the variability of the estimator. In a simple random sample, the degrees of freedom of a variance estimate depend upon the number of subjects and on the distribution of the variable under consideration. In the NAEP design, the degrees of freedom are primarily a function of the number of primary sampling units and the number of strata in the design, rather than the number of subjects, and the distribution of the variable under consideration has less impact. Section 10.4 discusses the degrees of freedom for NAEP jackknife variance estimates.

Since the sample design determines the derivation of the sampling weights and the estimation of sampling variability, it will be helpful to note the key features of the 1990 NAEP sample design. A description of the design appears in Chapter 3, and the various assessment instruments are detailed in Chapter 4.

The 1990 sample was a multistage probability sample consisting of four stages of selection. The first stage of selection, the primary sampling units (PSUs), consisted of counties or groups of counties. The second stage of selection consisted of elementary and secondary schools. The assignment of sessions to sampled schools comprised the third stage of sampling, and the fourth stage involved the selection of students within schools and their assignment to sessions. The probabilities of selection of the first-stage sampling units were proportional to measures of their size, while the probabilities for subsequent stages of selection were such that the overall probabilities of selection of students were approximately uniform, with exceptions for certain subpopulations that were oversampled by design. For the main assessment, schools with relatively high concentrations of Black and/or Hispanic students were deliberately sampled at twice the normal rate to obtain larger samples of respondents from those subpopulations, in order to increase the precision in the estimation of the characteristics of these subpopulations. Also for the main assessment, nonpublic school students were sampled at three times the normal rate, again so as to increase the precision of estimates for this population subgroup. For all assessment components, students from schools with smaller numbers of eligibles received lower probabilities of selection, as a means of enhancing the cost efficiency of the sample.

The 1990 main assessment includes three student cohorts: students who were *either* in the fourth grade *or* 9 years old; students who were *either* in the eighth grade *or* 13 years old; and students who were *either* in the twelfth grade *or* 17 years old. The main assessment represents two overlapping samples. The first sample represents students of specified grades (who could be of any age). The second sample represents students of specified ages (who could be of any grade). Students were age-eligible if they were born in the appropriate calendar year (1980,

1976, or 1972). The main assessment of all age/grade levels was conducted in the winter and spring of 1990 and the sample design was such that the students assessed in the winter and the students assessed in the spring constitute two representative samples of the population.

The full 1990 assessment also includes a number of additional samples which used the age definitions, times of testing, and modes of administration used in previous assessments. Because the purpose of these studies was to provide the statistical linkage between the 1990 data and data from previous assessments, they are referred to as bridge (or trend) studies.

The full 1990 NAEP assessment thus includes a number of different samples from several populations. Each of these samples has its own set of weights that are to be used to produce estimates about the characteristics of the population addressed by the sample (the target population). The various samples and their target populations are as follows:

The Main Samples of Students. These samples, one for each of the three age/grade levels, were drawn in the winter and spring, use the calendar-year age definitions, and consist of all students assessed in the main assessment. The target population for each of these samples consists of all students who are in the specified age/grade who were deemed assessable by their school.

For each grade for each season, there were four distinct assessments conducted in three distinct session types. The reading and science sessions were administered together in common sessions, and all sampling procedures were directed at drawing a single sample of students for this combined assessment. The assignment of individual students to a reading or science assessment respectively was achieved through spiraling of the booklets (see below and Chapter 4). The main mathematics assessment was administered separately, and thus used a distinct subsample of students. This also applied to the special assessment in mathematics, which was administered using a paced audiotape (see below).

Because of these administrative arrangements, the weighting procedures were applied separately to three distinct subsets of the main sample of assessed students at each age/grade for each season. A set of weights was produced corresponding to each assessment type. Thus in total there were 18 sets of final student weights for the sample of assessed students in the main assessment—one for each combination of three age/grades, two seasons, and three assessment types.

Bridge to 1984. This bridge consists of trend samples comparable to the 1984 main assessment and addresses the subject areas of reading and writing. The samples were collected by grade and age for age 9/grade 4, age 13/grade 8, and age 17/grade 11, using the age definitions and time of testing from 1984. Six assessment booklets were administered at each age/grade. The respondents to the combined set of assigned booklets at a given age/grade constitute a representative sample of the population of students who are of the specified grade *or* of the specified age. The respondents to any one of the booklets also constitute a representative sample.

Bridge to 1986, Ages 9 and 13. This bridge consists of samples for ages 9 and 13 comparable to those used for the measurement of trends in 1986. The samples were collected by age only and used the same age definitions and time of testing as in 1984 and in the 1986 bridge to 1984. Three assessment booklets were administered to each age group and the respondents to any one of the booklets assigned to a given age constitute a representative sample of the population of all students of that age. The subjects of mathematics, science, and reading were assessed.

Bridges to 1986, Grade 11/Age 17. A total of eight booklets were used at this age/grade to measure trends to 1986. There were two distinct assessment components, which represented somewhat different target populations and were weighted separately. One component, consisting of six booklets, assessed the subject of reading, mathematics and science for the population of students in eleventh grade or 17 years old. The second component involved two booklets, administered in distinct sessions but analyzed as a single sample, and assessed mathematics and science in the population of 17-year-old students.

Bridge for Long-term Trend. At each age, two booklets were administered to assess long-term trend for 9-, 13-, and 17-year-old students, using the age definitions appropriate in 1984 and earlier assessments (and the 1986 bridge assessments). The booklets were administered as separate sessions in each case but were weighted together as a single sample for each age. The subjects assessed were mathematics and science.

For purposes of sampling and weighting, the assessment samples are categorized as "tape" or "spiral" according to whether or not paced audiotapes are to be used in the administration:

1) *Tape samples* are samples that require audiotape pacing in the assessment (some of the bridges to 1986 and the special mathematics assessment). For these samples, all students within a particular assessment session receive the same booklet and are paced through at least part of the booklet with an audiotape. These assessment sessions are accordingly referred to as tape sessions.

2) *Spiral samples* are all main assessment samples and the remaining bridge samples. For these samples, no audiotape pacing was employed and the assessment booklets presented to a particular sample are spiraled through each assessment session (that is, the booklets are systematically interspersed and assigned for testing in that order). These assessment sessions are referred to as spiral sessions.

10.1 DERIVATION OF THE SAMPLE WEIGHTS

As indicated previously, NAEP uses differential sampling rates, deliberately oversampling certain subpopulations to obtain larger samples of respondents from those subgroups, thereby enhancing the precision of estimates of characteristics of these oversampled subgroups. As a result of the oversampling, these subpopulations, corresponding to students from schools with

high concentrations of Black and/or Hispanic students, and from nonpublic schools, are overrepresented in the sample. Lower sampling rates were introduced also for very small schools (those schools with only 1 to 19 eligible students). This reduced level of sampling from small schools was undertaken in an approximately optimum manner as a means of reducing variances per unit of cost. Appropriate estimation of population characteristics must take disproportionate representation into account. This is accomplished by assigning a weight to each respondent, where the weights approximately account for the sample design and reflect the appropriate proportional representation of the various types of individuals in the population.

The weighting procedures for 1990 included computing the student's base weight, the reciprocal of the probability that the student was invited to a particular session. Such weights are those appropriate for deriving estimates from probability samples via the standard Horvitz-Thompson estimator (see Cochran, 1977, section 9A.7). These base weights were adjusted for nonresponse and then subjected to a trimming algorithm to reduce a few excessively large weights. The weights were further adjusted by a poststratification procedure in an effort to reduce the sampling error and certain potential biases of estimates relating to student populations corresponding to several subgroups of the total population. Poststratification was performed by adjusting the weights of the sampled students so that the resulting estimates of the total number of students in a number of specified subgroups of the population corresponded to population totals based on information from the Current Population Survey and Census Bureau estimates of the population. The subpopulations were defined in terms of race, ethnicity, geographic region, age, and grade.

The following sections provide an overview of the procedures used to derive the sampling weights. Further details in the derivation of these weights can be found in *1990 National Assessment of Educational Progress Sampling and Weighting Procedures, Part 2: National Assessment, Final Report* (Rust, Burke, & Fahimi, 1992).

10.1.1 Student Base Weight

The base weight assigned to a student is the reciprocal of the probability that the student was invited to a particular type of assessment session. That probability is the product of four factors:

1) the probability that the PSU was selected;

2) the conditional probability, given the PSU, that the school was selected and (in the case of the main assessments) assigned to the specific season (winter or spring);

3) the conditional probability, given the sample of schools in a PSU, that the school was allocated the specified type of session; and

4) the conditional probability, given the school, that the student was invited to the specified type of session.

Thus, the base weight for a student may be expressed as the product

$$W_B = PSUWT \bullet SCHWT \bullet SESSWT \bullet STUSCHW$$

where PSUWT, SCHWT, SESSWT, and STUSCHW are, respectively, the reciprocals of the preceding probabilities.

Each school selected for the main assessment from within the sampled PSUs had a probability of 0.5 of being allocated to winter or spring. This was achieved by assigning alternate schools within each PSU to winter and spring, where the order used was the order in which the systematic sample was drawn within each PSU (see Chapter 3).

The school selection probability was modified to account for the fact that schools included in the sample for the 1990 National Educational Longitudinal Study (NELS) First Phase Follow-up were excluded from the NAEP assessment. To make this adjustment, the probability of selection for each school in the sample, derived from the systematic sampling scheme, was multiplied by an estimate of the probability that the school was not included in the NELS sample. This probability could not be calculated in a straightforward manner since the NELS sample of schools consisted of those attended by a probability sample of students, drawn in 1988 when the students were in the eighth grade. For schools having both grades 8 and 10, and included in a NAEP sample, we assumed that the probability of inclusion in the 1990 NELS sample was equal to the probability of inclusion of the school in the 1988 sample (that is, that the selected students who were being followed in the study had in the main remained within the same school in such a case). For schools having grade 10 but not grade 8, a different procedure was used to estimate the probability that the school was included in the NELS sample. As part of the data collection at each such school, a list of names was obtained of schools from which eighth graders eventually graduated to attend tenth grade in the NAEP sampled school. The probability of selection of the school in the 1988 NELS school sample was determined from the NELS frame, and these probabilities were combined across all of the schools listed by a given NAEP sample school to give an estimate of the probability that at least one of the schools was included in the 1988 sample. This probability estimate was taken as the probability of inclusion in the 1990 NELS school sample, based on two assumptions. The first assumption was that, had any of the listed schools been included in the 1988 NELS sample, then the subject school would have been included in the 1990 NELS sample, since students would have been followed there from the selected listed school. The second assumption was that the chances of selection for the listed schools were independent, which permitted the calculation of the probability that at least one such school was selected.

These base weight adjustments primarily affected the age 17/grade 11 and age 17/grade 12 samples, since it was predominantly schools in these samples that had grade 10.

Tables C-1 to C-3 in Appendix C show the distribution of base weights for each of the separate sessions conducted as part of the 1990 assessment. The variations in probabilities of selection, and consequently of weights, were introduced by design, either to increase the effectiveness of the sample in achieving its goals of reporting for various subpopulations, or to achieve increased efficiency per unit of cost.

10.1.2 Adjustment of Base Weights for Nonresponse

The base weight for a selected student was adjusted by five nonresponse factors. The first of these was to adjust for noncooperating schools, while the second was needed to adjust for allocated sessions that occasionally were not conducted. The third adjustment was needed to account for those few cases where, either inadvertently or on the insistence of the school, only students in the modal grade were given a chance of inclusion in the sample. The fourth adjustment conversely was needed to account for those few cases where, inadvertently, only students of the modal age were given the chance of inclusion in the sample. By definition this adjustment was never required for the bridge tape samples, for which only modal age students were invited. The fifth adjustment was needed to adjust for students who were (or should have been) invited to the assessment but did not appear either in the scheduled session or a makeup session. Thus, the nonresponse adjusted weight for a student is of the form

$$W' = W_B \bullet SCHNRF \bullet SESNRF \bullet AOENRF \bullet GOENRF \bullet STUNRF$$

where the nonresponse adjustment factors SCHNRF, SESNRF, AOENRF, GOENRF, and STUNRF are computed as described below. It should be noted that the nonresponse adjustments assume that nonresponse occurs at random within the categories within which adjustments are made. Some degree of bias could results to the extent that this assumption is false.

10.1.2.1 School Nonresponse Adjustment (SCHNRF)

The school nonresponse adjustment was intended to compensate for school nonresponse occurring before session assignment. These factors were computed separately within a subuniverse (see Chapter 3); that is within one of 18 classes of PSUs within the same region, certainty/noncertainty status, MSA/nonMSA status, and high minority status.

The school nonresponse adjustment factor in subuniverse h, $SCHNRF_h$, is given by

$$SCHNRF_h = \frac{\sum_{i \in A_h} PSUWT_{hi} \cdot SCHWT_{hi} \cdot G_{hi}}{\sum_{i \in B_h} PSUWT_{hi} \cdot SCHWT_{hi} \cdot G_{hi}}$$

where

$PSUWT_{hi}$ = the PSU weight for the PSU containing school i from subuniverse h

$SCHWT_{hi}$ = the school weight for school i in subuniverse h;

G_{hi} = the estimated number of age/grade-eligible students in school i in subuniverse h based on QED data (for sessions involving only age-eligible students, the number of age eligibles in each school was used);

set A_h consists of the *original* sample of schools (cooperating and noncooperating schools, but not substitutes); and

set B_h . consists of all schools cooperating at the time of session allocation (including schools that were substituted for noncooperating schools).

For a substitute school, $SCHWT_{hi}$ is defined as the school weight of the originally selected school, while the value of G is taken from the substitute school itself. In those cases where PSUs were combined, the value of PSUWT was included in the numerator and denominator of the school nonresponse adjustment factor. Occasionally two subuniverses were combined to form a single nonresponse class. This occurred when the number of participating schools from within a subuniverse was small, leading to un lue instability in the school nonresponse adjustment factor prior to such collapsing. Subuniverses collapsed together were as similar in nature as possible.

Tables C-4 to C-6 in Appendix C show the distribution of school nonresponse adjustment factors for each of the 1990 assessment sessions.

10.1.2.2 Session Nonresponse Adjustment (SESNRF)

The session nonresponse adjustment was intended to compensate for school nonresponse occurring after session assignment. These factors were computed separately within the same nonresponse classes as were used for school of nonresponse adjustment, except that occasionally additional collapsing of classes was necessary, especially for the smaller assessment components.

In PSU h, the session nonresponse adjustment factor $SESNRF_h$ was given by

$$SESNRF_h = \frac{\sum_{i \in B'_h} PSUWT_{hi} \cdot SCHWT_{hi} \cdot SCHNRF_{hi} \cdot SESSWT_{hi} \cdot G_{hi}}{\sum_{i \in C_h} PSUWT_{hi} \cdot SCHWT_{hi} \cdot SCHNRF_{hi} \cdot SESSWT_{hi} \cdot G_{hi}}$$

where

$PSUWT_{hi}$ = the PSU weight for the PSU containing school i frc subuniverse h;

$SCHWT_{hi}$ = the school weight for school i in subuniverse h;

$SCHNRF_{hi}$ = the school nonresponse adjustment for school i in subuniverse h;

$SESSWT_{hi}$ = the session allocation weight for school i in subuniverse h;

G_{hi} = the estimated number of age/grade-eligible students in school i in subuniverse h in the case of spiral sessions, and the estimated number of age-eligible students in the case of the tape sessions, to which only age eligibles were invited (the values of G_{hi} were based on QED data);

set B'_h consists of all in-scope schools allocated to the specific type of session in subuniverse h that were to be participating at the time of session allocation; and

set C_h consists of all schools allocated to the session type in subuniverse h that ultimately participated.

Tables C-7 to C-9 in Appendix C show the distribution of the session nonresponse adjustment factor for each of the 1990 sample sessions.

10.1.2.3 Age-only Eligibles Nonresponse Adjustment (AOENRF)

In a few schools in which assessments took place, only those students in the modal grade were listed for sampling (see Chapter 3), even though there was definite or very strong evidence that other age-eligible students were enrolled. Thus, an adjustment factor was needed to account for the fact that, although students eligible by age alone (age-only eligibles) were almost certainly enrolled in these schools, they were not given a chance of inclusion in the sample. These factors were calculated separately by subuniverse.

The school-level age-only eligibles nonresponse adjustment factor in subuniverse h, $AOENRF_h$, is given for students not in the modal grade by

$$AOENRF_h = \frac{\sum_{i \in D_h} PSUWT_{hi} \cdot SCHWT_{hi} \cdot SCHNRF_{hi} \cdot SESSWT_{hi} \cdot SESNRF_{hi} \cdot AO_{hi}}{\sum_{i \in C_h} PSUWT_{hi} \cdot SCHWT_{hi} \cdot SCHNRF_{hi} \cdot SESSWT_{hi} \cdot SESNRF_{hi} \cdot AO_{hi}}$$

where

$PSUWT_{hi}$ = the PSU weight for the PSU containing school i from subuniverse h;

$SCHWT_{hi}$ = the school weight for school i in subuniverse h;

$SCHNRF_{hi}$ = the school nonresponse adjustment for school i in subuniverse h;

$SESSWT_{hi}$ = the session allocation weight for school i in subuniverse h;

$SESNRF_{hi}$ = the session nonresponse adjustment for school i in subuniverse h;

AO_{hi} = the estimated number of age-only eligible students in school i, subuniverse h, based on data from the principal's questionnaire;

set C_h consists of all schools allocated to the particular session type in subuniverse h that ultimately participated; and

set D_h consists of all schools allocated to the particular session type in subuniverse h, that could be reasonably supposed to have included age-only eligible students in the assessment, if any.

The value of $AOENRF_h$ for students in the modal grade is given as 1.0, since they were not subject to this component of nonresponse.

10.1.2.4 Grade-only Eligibles Nonresponse Adjustment (GOENRF)

In a very few schools in which assessments took place, for sessions for which both grade- and age-eligible students were to be invited, only students who were eligible by age were invited to the assessment. This was of course appropriate for those schools that did not have the modal grade, and was also probably appropriate in a number of small schools where all students in the modal grade were age-eligible. In the case of number of these schools, however, there was strong evidence that the school had enrolled students who were eligible for the assessment by grade alone, but who were not given a chance of inclusion in the assessment sample. Thus an adjustment factor was needed to account for this fact. These factors were calculated separately by subuniverse.

The school-level grade-only eligibles nonresponse adjustment factor in subuniverse h, $GOENRF_h$, is given for students not of modal age by

$$GOENRF_h = \frac{\sum_{i \in E_h} PSUWT_{hi} \cdot SCHWT_{hi} \cdot SCHNRF_{hi} \cdot SESSWT_{hi} \cdot SESNRF_{hi} \cdot GO_{hi}}{\sum_{i \in C_h} PSUWT_{hi} \cdot SCHWT_{hi} \cdot SCHNRF_{hi} \cdot SESSWT_{hi} \cdot SESNRF_{hi} \cdot GO_{hi}}$$

where

$PSUWT_{hi}$ = the PSU weight for the PSU containing school i in subuniverse h;

$SCHWT_{hi}$ = the school weight for school i in subuniverse h;

$SCHNRF_{hi}$ = the school nonresponse adjustment for school i in subuniverse h;

$SESSWT_{hi}$ = the session allocation weight for school i in subuniverse h;

$SESNRF_{hi}$ = the session nonresponse adjustment for school i in subuniverse h;

GO_{hi} = the estimated number of grade-only eligible students in school i, subuniverse h, based on data from the principal's questionnaire;

set C_h consists of all schools allocated to the particular session type in subuniverse h that ultimately participated; and

set E_h consists of all schools allocated to the particular session type in subuniverse h, that could be reasonably supposed to have included grade-only eligible students in the assessment, if any.

The value of $GOENRF_h$ for students in the modal age is given as 1.0, since they were not subject to this component of nonresponse.

For a given student, either AOENRF or GOENRF must equal 1.0. Tables C-10 to C-12 in Appendix C show the distribution of the product of these two nonresponse adjustments for each of the 1990 assessment sessions.

10.1.2.5 Student Nonresponse Adjustment (STUNRF)

Student nonresponse adjustment factors were completed separately for each of the assessment session types within each PSU.

For spiral sessions, the student nonresponse adjustment was made separately for two classes of students in PSU: those in or above the modal grade for their age, and those below. This differentiation acknowledges likely differences between students in the two classes, both in their assessed abilities and in their likelihood of nonresponse. For some sessions in some PSUs, these two classes were combined, since one or both was too small to form the basis for an adjustment factor. For each class c in PSU h, the student nonresponse adjustment factor $STUNRF_{hc}$ is computed by

$$STUNRF_{hc} = \frac{\sum_{j(l) \in A'_{hc}} SCHWT_{hi} \cdot SCHNRF_{hi} \cdot SESSWT_{hi} \cdot SESNRF_{hi} \cdot AOENRF_{hi} \cdot GOENRF_{hi} \cdot STUSCHW_{hij(l)}}{\sum_{j(l) \in B'_{hc}} SCHWT_{hi} \cdot SCHNRF_{hi} \cdot SESSWT_{hi} \cdot SESNRF_{hi} \cdot AOENRF_{hi} \cdot GOENRF_{hi} \cdot STUSCHW_{hij(l)}}$$

where

$SCHWT_{hi}$ = the school weight for school i in PSU h;

$SCHNRF_{hi}$ = the school nonresponse adjustment factor for school i in PSU h;

$SESSWT_{hi}$ = the session allocation weight for the particular session in school i in PSU h;

$SESNRF_{hi}$ = the session nonresponse adjustment factor in school i in PSU h;

$AOENRF_{hi}$ = the age-only eligibles nonresponse adjustment factor for school i in PSU h;

$GOENRF_{hi}$ = the grade-only eligibles nonresponse adjustment factor for school i in PSU h;

$STUSCHW_{hij}$ = the within-school student weight for student j in school i in PSU h;

Set A'_{hc} consists of the students in class c in PSU h who were invited to the session; and

Set B'_{hc} consists of the students in class c in PSU h who were assessed in the session.

The student nonresponse adjustment for tape sessions was similar, except that the adjustment was computed within a PSU for each tape session type across all students originally invited to the assessment for that session type. This was consistent with past practice for tape sessions.

Tables C-13 to C-15 in Appendix C show the distribution of student nonresponse adjustment factors for each of the 1990 assessment sessions.

10.1.2.6 Evaluation of Potential for Bias Resulting from School and Student Nonparticipation

Although school and student nonresponse adjustments are intended to reduce the potential for nonparticipation to bias the assessment results, they cannot completely eliminate this potential bias with certainty. The extent of bias remains unknown, of course, since there are not assessment data for the nonparticipating schools and students.

Some insight can be gained about the potential for residual nonresponse bias, however, by examining the weighted school- and student-level distributions of characteristics known for both participants and nonparticipants, especially for those characteristics known or thought likely to be related to achievement on the assessment. If the distributions for the full sample of schools (or students) without the use of nonresponse adjustments are close to those for the participants with nonresponse adjustments applied, there is reason to be confident that the bias from nonparticipation is small.

There are no suitable student-level characteristics readily available for both absent and assessed students, but there are several school-level characteristics available for both assessed and absent students, and for participating and nonparticipating schools. The tables below show the combined impact of nonresponse and of the nonresponse adjustments on the distributions of schools (weighted by the estimated number of eligible students enrolled) and students, by the type of school (public, Catholic, other private) and the size of the school, as measured by the estimated number of eligible students enrolled. Three size classes have been defined for each age class. For school nonresponse, we have also considered the urban/rural nature of the county. The data are for the main assessments—all session types combined and winter and spring assessments combined.

Table 10-1 shows the weighted marginal distributions of schools for each of the three classification variables for each age class, using the full sample of in-scope schools—those participating, plus those refusals for which no substitute participated. Table 10-2 shows the same distribution based only on participating schools, with school nonresponse adjustments applied to them. For school-level data, the school nonresponse adjustment is actually a

Table 10-1

Distribution of Populations of Eligible Students
Based on Full Weighted Sample of Eligible Schools
1990 Main NAEP Samples

	Age 9/Grade 4	Age 13/Grade 8	Age 17/Grade 12
Total population	4,667,667	4,666,540	4,421,666
School Type			
Catholic	6.42%	6.56%	4.29%
Other Private	5.49%	4.14%	4.70%
Public	88.09%	89.29%	91.01%
School Size*			
1	16.06%	10.90%	10.59%
2	34.39%	42.96%	42.48%
3	49.55%	46.14%	46.92%
County Type (SDOC)**			
Central city, 200,000+	34.22%	31.93%	27.24%
Other, 200,000+	14.31%	16.62%	18.86%
Other 25,000+	22.80%	20.52%	26.31%
Other	23.37%	25.50%	20.45%
Extreme Rural	5.30%	5.43%	7.14%

* School size = number of eligible students enrolled:

	1	2	3
Age 9/Grade 4	1-49	50-99	100+
Age 13/Grade 8	1-49	50-299	300+
Age 17/Grade 12	1-99	100-399	400+

** County type is given by the county-level variable "Sampling Description of Community" (SDOC). For a full definition, see Rust et al. (1992).

Table 10-2

Distribution of Populations of Eligible Students
Based on Weighted Sample of Participating Schools with School Nonresponse Adjustments
1990 Main NAEP Samples

	Age 9/Grade 4	Age 13/Grade 8	Age 17/Grade 12
Total population	4,667,667	4,666,540	4,421,666
School Type			
Catholic	6.74%	7.63%	4.75%
Other Private	5.10%	3.63%	3.69%
Public	88.16%	88.73%	91.56%
School Size*			
1	16.27%	11.37%	10.02%
2	33.96%	41.52%	40.91%
3	49.76%	47.11%	49.08%
County Type (SDOC)**			
Central city, 200,000+	33.64%	32.04%	26.90%
Other, 200,000+	14.88%	17.11%	20.13%
Other 25,000+	23.21%	20.31%	25.62%
Other	23.26%	25.34%	20.08%
Extreme Rural	5.01%	5.21%	7.27%

* School size = number of eligible students enrolled:

	1	2	3
Age 9/Grade 4	1-49	50-99	100+
Age 13/Grade 8	1-49	50-299	300+
Age 17/Grade 12	1-99	100-399	400+

** County type is given by the county-level variable "Sampling Description of Community" (SDOC). For a full definition, see Rust et al. (1992).

composite of the school and session nonresponse adjustment factors derived for use with student-level data (see section 10.1.6).

It can be seen from the tables that, even though the level of school nonparticipation is as high as 18.7 percent for age class 17 (see Table 3-9), and somewhat lower for the other age classes (13.3 percent for age class 13 and 11.4 percent for age class 9), the distributions for the three characteristics considered remain similar.

Tables 10-3 and 10-4 present similar data for students. Table 10-3 shows the distributions for assessed and absent students (with base weights adjusted for school nonparticipation) while Tables 10-4 shows them for assessed students only, with the student nonresponse adjustments also applied to the weights. The rates of student nonparticipation were 7.1 percent for age class 9, 10.9 percent for age class 13, and 18.7 percent for age class 17 (see Table 3-9). The tables show that for the distributions of type of school attended and size of school attended, the combined effect of student nonparticipation and the subsequent nonresponse adjustments have resulted in very little change in distribution.

10.1.3 Trimming of Weights

In a number of cases, students were assigned relatively large weights. One cause of large weights was underestimation of the number of eligible students in some schools leading to inappropriately low probabilities of selection for those schools. A second major cause is the presence of large schools (high schools in particular) in PSUs with small selection probabilities. In such cases, the maximum permissible within-school sampling rate (determined by the maximum sample size allowed per school—see Chapter 3) could well be smaller than the desired overall within-PSU sampling rate for students. Large weights arose also because very small schools were, by design, sampled with low probabilities. Other large weights arose as the result of high levels of nonresponse coupled with low to moderate probabilities of selection, and the compounding of nonresponse adjustments at various levels.

Students with notably large weights have an unusually large impact on estimates such as weighted means. Since, under some simplifying assumptions, the variability in weights contributes to the variance of an overall estimate by an approximate factor $1 + V^2$, where V^2 is the relative variance of the weights, an occasional unusually large weight is likely to produce large sampling variances of the statistics of interest, especially when the large weights are associated with students with atypical performance characteristics.

To reduce this problem, a procedure of trimming a few of the more extreme weights to values somewhat closer to the mean weight was applied. This trimming can increase the accuracy of the resulting survey estimates, substantially reducing V^2 and hence the sampling variance while introducing a small bias. The trimming algorithm was identical to that used in the 1984, 1986, and 1988 assessments and had the effect of trimming the weights of students from any school that contributed more than a specified proportion, ζ, to the estimated variance of the estimated number of students eligible for assessment. The trimming was done separately for each assessment session type. In each case, the value of the proportion ζ was chosen to be $10/K$, where K was the number of schools in which a specified assessment was conducted. The

Table 10-3

Distribution of Populations of Eligible Students
Based on Assessed and Absent Students from Participating Schools
1990 Main NAEP Samples

	Age 9/Grade 4	Age 13/Grade 8	Age 17/Grade 12
Total population	4,641,929	4,231,363	3,629,066
School Type			
Catholic	6.09%	7.16%	4.91%
Other Private	4.12%	3.38%	4.06%
Public	89.79%	89.46%	91.03%
School Size*			
1	16.00%	11.14%	11.06%
2	35.20%	42.92%	42.59%
3	48.80%	45.95%	46.36%

Table 10-4

Distribution of Populations of Eligible Students
Based on Assessed Students from Participating Schools with Student Nonresponse Adjustments Applied
1990 Main NAEP Samples

	Age 9/Grade 4	Age 13/Grade 8	Age 17/Grade 12
Total population	4,641,929	4,231,363	3,629,066
School Type			
Catholic	6.19%	6.92%	4.60%
Other Private	4.13%	3.30%	3.86%
Public	89.68%	89.78%	91.54%
School Size*			
1	16.09%	10.98%	10.84%
2	35.38%	43.12%	42.43%
3	48.54%	45.90%	46.73%

* School size = number of eligible students enrolled:

	1	2	3
Age 9/Grade 4	1-49	50-99	100+
Age 13/Grade 8	1-49	50-299	300+
Age 17/Grade 12	1-99	100-399	400+

number of schools where weights were trimmed was small in each of the samples. Tables C-16 to C-18 in Appendix C show the distribution of trimming factors for each of the 1990 assessment sessions. From the table it is seen that the most extreme trimming factors applied were of the order of 0.5 and that trimming affects the weights of only a very small proportion of the assessed students.

10.1.4 Poststratification

As in most sample surveys, the respondent weights are random variables that are subject to sampling variability. Even if there were no nonresponse, the respondent weights would at best provide unbiased estimates of the various subgroup proportions. However, since unbiasedness refers to average performance over a conceptually infinite number of replications of the sampling, it is unlikely that any given estimate, based on the achieved sample, will exactly equal the population value. Furthermore, the respondent weights have been adjusted for nonresponse and a few extreme weights have been reduced in size.

To reduce the mean squared error of estimates using the sampling weights, these weights were further adjusted so that estimated population totals for a number of specified subgroups of the population, based on the sum of weights of students of the specified type, were the same as presumably better estimates based on composites of estimates from the 1987 and 1988 Current Population Survey and 1990 population projections made by the Census Bureau. This adjustment, called poststratification, is intended especially to reduce the mean squared error of estimates relating to student populations that span several subgroups of the population, and thus to reduce the variance of measures of changes over time for such student populations.

10.1.4.1 1990 Poststratification Procedures

The poststratification in 1990 was done for each age/grade and separately for each of the spiral assessments and each of the groups of tape assessments. Within each age/grade and assessment type group, poststratification adjustment cells were defined in terms of race, ethnicity, and NAEP region as shown in Table 10-5.

The result is seven poststratification cells for each group of tape sessions. For the assessments involving both age and grade eligible students, each of the seven subgroups was further divided into two or three eligibility classes. For age classes 9 and 13 and for the age 17/grade 11 bridge sample, three eligibility classes were used:

a) students eligible by grade and of modal age;
b) students eligible by age only;
c) students eligible by grade but not of modal age.

For the age 17/grade 12 main assessment sample, each of the seven subgroups was divided into two subclasses:

a) students eligible by grade (of any age);
b) students eligible by age only.

This variation in the procedure from that used for the other age classes and for the age 17/grade 11 bridge was adopted because the independent estimates of the numbers of students in the population did not provide consistent data on the numbers of twelfth grade students eligible only by grade. This procedure for age 17/grade 12 is identical to that employed for the 1988 assessment. (See Rust, Bethel, Burke, & Hansen, 1990, for further details.)

Table 10-5
Major Subgroups for Poststratification in 1990

Subgroup	Race	Ethnicity	Region*
1	White	Not Hispanic	Northeast
2	White	Not Hispanic	Southeast
3	White	Not Hispanic	Central
4	White	Not Hispanic	West
5	Any	Hispanic	Any
6	Black	Not Hispanic	Any
7	Other	Not Hispanic	Any

* Regions are the same as for stratification and reporting (see Chapter 3), except that all of Virginia is included in the Southeast region for poststratification purposes.

Thus, there were 7, 14, or 21 cells for poststratification. The poststratified weight for each student within a particular cell was the student's base weight, with adjustments for nonresponse and trimming, times a poststratification factor. For each cell, the poststratification factor is a ratio whose denominator is the sum of the weights (after adjustments for nonresponse and trimming) of assessed and excluded students, and whose numerator is an adjusted estimate of the total number of students in the population who are members of the cell. This estimated total was a composite based on the October 1987 and 1988 Current Population Surveys and 1990 population projections. Details of the procedures used to obtain these composite independent estimates are provided in Rust et al. (1992). Tables C-19 to C-21 in Appendix C show the distribution of poststratification factors for each of the 1990 assessments.

10.1.4.2 Differences from Earlier Procedures

The poststratification procedure utilized for 1990 was similar to that used in 1988. This differed somewhat from the procedures used in 1986 and 1984, and the nature and impact of these differences are discussed in Johnson and Zwick (1990, section 10.1.4). There were two differences from the 1988 procedures. The first was in the definitions of the samples that were to be poststratified to the appropriate population totals for the assessments involving only age eligible students. In 1988 each individual bridge tape session was separately poststratified (these are the session types where only age eligible students are assessed) whereas in 1990 these sessions were formed into groups of sessions which essentially constituted a single assessment (two groups at each age), and these groups were poststratified (and indeed weighted in general) as a whole. Conversely, as in 1988 the spiral sessions were poststratified individually, although in most cases a number of different booklets were administered in a single session.

This change from 1988 was designed to provide a consistent approach whereby all the assessment components that, although administered distinctly, were analyzed together were poststratified together to the total population. This avoided the potential problems of increased variance that arise when small individual session types are poststratified, with resulting high sampling variability in the poststratification factors derived.

The second difference affected those assessment components to which both age- and grade-eligible students were invited, for age 9/grade 4 and age 13/grade 8. In both the 1988 and the 1990 assessments, students in the modal grade were classified into two groups on the basis of age. In 1988 this split was determined by whether or not the student was age eligible (in addition to being grade eligible). In 1990 the split was determined by whether or not the student was of the modal age (in years) for the grade as of October 1, 1989. This change was made because the independent estimates of these two components of the population for each grade could be made more reliably under the 1990 procedure. This is because the 1990 classification was consistent with the data available from the Current Population Survey estimates, whereas in 1988 a modification was needed, which was based on certain assumptions about the joint distribution of the population by grade and age. These assumptions were in general supported by the available data, but were of necessity simplifying in nature, and therefore a potential source for a small amount of error. Full details of the 1988 procedure for obtaining the independent estimates are given in Appendix C of Rust et al. (1990).

For the age 17/grade 11 bridge spiral sessions and the age 17/grade 12 bridge samples, no changes were made from the 1988 procedures. At age 17/grade 11, the definition of age eligibility coincided with the modal age as of October 1, 1989, so that there was no change necessary. For age 17/grade 12, there was no split of the grade 12 students by age for purposes of poststratification, either in 1988 or 1990, as discussed above and in Appendix D of Rust et al. (1990).

10.1.5 The Final Student Weight

The final weight assigned to a student is the student full-sample weight. This weight is the student's base weight after the application of the various adjustments described above. The student full-sample weight was used to derive all estimates of population and subpopulation characteristics that have been presented in the various NAEP reports, including simple estimates such as the proportion of students of a specified type who would respond in a certain way to an item and more complex estimates such as mean proficiency levels.

The effects of all of the adjustments to the base weights are summarized in Tables C-22 to C-24 in Appendix C, which show the distribution of the single factor given as the product of SCHNRF, SESNRF, AOENRF, GOENRF, STUNRF, the trimming factor, and the poststratification factor, for each of the assessment components. The distributions of the final student weights are given in Tables C-25 to C-27 in Appendix C.

As indicated earlier, under some simplifying assumptions the factor $1 + V^2$ indicates the approximate relative increase in variance of estimates resulting from the variability in the weights. The factor $1 + V^2$ for each sample is readily derivable from Tables C-25 to C-27 by adding 1 to the square of the ratio of the standard deviation to the mean weight. These factors,

resulting from the combined effect of the variations in weights introduced by design and from other causes, are discussed above.

10.1.6 Other Weights

In addition to the weights for the assessed students, weights were also derived for excluded students. In addition, a special weighting adjustment was developed for a subset of the eighth-grade students assessed in the main sample spiral mathematics assessment. The weights for this subset, with this adjustment applied, were used in equating the results of the national eighth-grade mathematics assessment and the Trial State Assessment. Finally, a set of weights appropriate for analyzing school-level data files was developed.

Weights for excluded students. Excluded students are students who were designated by the schools as unable to complete the assessment because they were non-English speaking, mildly mentally retarded (educable), or functionally disabled. Since the same grade and age eligibility definitions apply, no distinction is made between students excluded from the various sessions within an assessment. However, the excluded students from the bridge assessments and from the winter and spring main assessments were treated as three separate samples of excluded students for each age class. This was in part because the guidelines to school personnel for excluding students differed between the main and bridge samples, so that the excluded student populations may have differed between these assessment types. The distribution of the final weights for excluded students and the components of the weights are included in Tables C-1 to C-27 in Appendix C.

For the bridge samples, students could be potentially excluded from a tape session for which only age eligible students were selected, or a spiral session, for which both grade and age eligible students were selected. The samples of excluded students from the bridge assessments were weighted to reflect the full grade- and/or age-eligible population. This was achieved by weighting each grade-only eligible (i.e., not eligible by age) student who was excluded from a spiral session to account for his/her probability of assignment to a spiral session. No such corresponding session assignment adjustment was needed for the age-eligible excluded students, since they were eligible to have been selected for any of the bridge assessment session types.

As in the case of the weights for the assessed students, the excluded student weights were constructed from components reflecting the probability of selection, correction for nonresponse, weight trimming, and poststratification. Further details on the derivation of the excluded student weights can be found in Rust et al. (1992).

Weights for equating national and state-by-state eighth-grade public-school mathematics assessments. The eighth-grade mathematics assessment conducted in February 1990 in each of 37 states, the District of Columbia, and two territories consisted of identical assessment material to that administered in the age 13/grade 8 main sample mathematics sessions. Technical details of the Trial State Assessment Program are given in Koffler (1991). The national and state-by-state assessments were equated so that state and national results could be reported on a common scale. The equating was achieved by using from each assessment that part of the

sample representing a common population. For the national sample, this consisted of those eighth-grade public-school students from a participating state (including the District of Columbia) who were assessed in the main sample mathematics assessment during the winter assessment period.

Although this sample of students received appropriate weights from the weighting procedure used for the national assessment, in an effort to increase the precision of the equating process, an additional weighting adjustment was developed and applied to this subsample, solely for use in equating. The adjustment involved adjusting the distributions of the weights for three categorical variables to agree closely with those obtained from the weighted aggregate sample from the state assessments from the participating states. The three variables were region (Northeast, Southeast, Central, and West); race/ethnicity (White nonHispanic, Black nonHispanic, Hispanic, and "other"), and type of mathematics course taken (algebra, pre-algebra, eighth-grade mathematics, and "other"). The equating of the weight distribution was achieved using a procedure known as Iterative Proportional Fitting (IPF), described in Bishop, Fienberg, and Holland (1975, Chapter 3). Ratio adjustments were applied to the national sample weights to force their distribution to agree with that from the aggregated state samples, for each of these three variables in turn. This process was then repeated, and the final set of adjusted weights was compared with the state sample weights on all three distributions, and found to be in very close agreement, with the maximum discrepancy across the 12 marginal cells being 0.9 percent. The resulting adjustments to the national weights ranged in magnitude from a factor of 0.704 (for White students from the Southeast taking the "other" mathematics course option) to 2.419 (for Hispanic students from the Northeast taking pre-algebra).

School weights. The sampling procedures used to obtain national probability samples of assessed students also gave rise indirectly to several national probability samples of schools (from which the students were subsequently sampled). So that the school samples can be utilized for making national estimates about schools, appropriate nonresponse adjusted survey weights have been developed.

The weight for each school is partly composed of a base weight, giving the inverse of the selection probability of the school. This weight, W_{BS}, is given by

$$W_{BS} = PSUWT \bullet SCHWT$$

School nonresponse adjustments were then applied to these base weights. These are very similar to the school nonresponse adjustment factors used for student weights, SCHNRF, and were created using the same set of nonresponse adjustment classes. The values of the adjustment factors are not the same, however. A school that was assigned a proper subset of the possible assessment sessions for a given assessment but did not participate at all was treated as not responding at the session level for the student weighting (since its nonparticipation did not affect those session types that were not assigned to it). Such a school was treated as a nonresponding school in creating the school weights.

A total of six samples of schools were weighted to be nationally representative. At each age/grade level, there were two such samples, one being the sample of schools selected for the bridge assessment, and the second being the sample of schools selected for the main assessment,

regardless of season (winter or spring). At age 9/grade 4, the population of schools represented in each case consists of all schools having at least one of the grades 2 through 5. The school population at age 13/grade 8 is that of schools having at least one of the grades 6 through 9, while the school population for both age 17/grade 11 and age 17/grade 12 is that of schools having at least one of the grades 9 through 12.

Jackknife replicate weights. In addition to the weights that were used to derive all estimates of population and subpopulation characteristics, other sets of weights, called jackknife replicate weights, were derived to facilitate the estimation of sampling variability by the jackknife variance estimation technique. These weights and the jackknife estimator are discussed in the next section.

10.2 PROCEDURES USED BY NAEP TO ESTIMATE SAMPLING VARIABILITY

A major source of uncertainty in the estimation of the value in the population of a variable of interest exists because information about the variable is obtained on only a sample from the population. To reflect this fact, it is important to attach to any statistic (e.g., a mean) an estimate of the sampling variability to be expected for that statistic. Estimates of sampling variability provide information about how much the value of a given statistic would be likely to change if the statistic had been based on another, equivalent, sample of individuals drawn in exactly the same manner as the achieved sample.

Another important source of variability is that due to imprecision in the measurement of individual proficiencies. For the 1990 assessment, proficiencies in subject areas except writing were summarized through item response theory (IRT) models, but not in the way that these models are used in standard applications where each person responds to enough items to allow for precise estimation of that person's proficiency. In NAEP, each individual responds to relatively few items so that individual proficiency values are not well determined. Consequently, the variance of any statistic based on proficiency values has a component due to the imprecision in the measurement of the proficiencies of the sampled individuals in addition to a component measuring sampling variability. The estimation of the component of variability due to measurement imprecision and its effect on the total variability of statistics based on proficiency values are discussed in Chapter 11.

The estimation of the sampling variability of any statistic must take into account the sample design. In particular, because of the effects of cluster selection (students within schools, schools within PSUs) and because of effects of nonresponse and poststratification adjustments, observations made on different students cannot be assumed to be independent of each other (and are, in fact, generally positively correlated). Furthermore, to account for the differential probabilities of selection (and the various adjustments), each student has an associated sampling weight, which should be used in the computation of any statistic and which is itself subject to sampling variability. Ignoring the special characteristics of the sample design and treating the data as if the observations were independent and identically distributed, will generally produce underestimates of the true sampling variability.

The prope : estimation of the sampling variability of a statistic based on the NAEP data is complicated and requires techniques beyond those commonly available in standard statistical packages. Fortunately, the *jackknife* procedure (see, e.g., Wolter, 1985; Kish & Frankel, 1974; Rust, 1985) provides good quality estimates of the sampling variability of most statistics, at the expense of increased computation, and can be used in concert with standard statistical packages to obtain a proper estimate of sampling variability.

The jackknife procedure used by NAEP has a number of properties that make it particularly suited for the analysis of NAEP data. When properly applied, a jackknife estimate of the variability of a linear estimator (such as a total) will be the same as the standard textbook variance estimate specified for the sample design (if the first-stage units were sampled with replacement and approximately so otherwise). Additionally, if the finite sampling corrections for the first stage units can be ignored, the jackknife produces asymptotically consistent variance estimates for statistics such as ratios, regression estimates or weighted means and for any other nonlinear statistic that can be expressed a smooth function of estimated totals of one or more variables (Krewski & Rao, 1981).

Through the creation of student replicate weights (defined below), the jackknife procedure allows the measurement of variability attributable to the use of poststratification and other weight adjustment factors that are dependent upon the observed sample data. Once these replicate weights are derived, it is a straightforward matter to obtain the jackknife variance estimate of any statistic.

The jackknife procedure in this application is based upon the development of a set of 56 jackknife replicate weights for each assessed student (or excluded student, or school depending upon the file involved). The 56 replicate weights are developed in such a way that, when utilized as described below, approximately unbiased estimates of the sampling variance of an estimate result, with an adequate number of degrees of freedom to be useful for purposes of making inferences about the parameter of interest. For a discussion of the degrees of freedom for variance estimation, see section 10.4.

The estimated sampling variance of a parameter estimator t is the sum of 56 squared differences:

$$\hat{Var}(t) = \sum_{i=1}^{56} (t_i - t)^2$$

where t_i denote the estimator of the parameter of interest, obtained using the ith set of replicate weights, $SRWT_i$, in place of the original set of full sample estimates WT_i. The methods for deriving these replicate weights, $SRWT_i$, are outlined below and full details are given in Rust et al. (1992).

Of the 56 replicate weights formed for each record, 30 act to reflect the amount of sampling variance contributed by the noncertainty strata of PSUs, with the remaining 26 replicate weights reflecting the variance contribution of the certainty PSU samples.

The derivation of the 30 replicate weights reflecting the variance of the noncertainty PSUs involves first defining pairs of PSUs (or appropriate aggregates of them in some strata) in a manner that models the design as one in which two PSUs are drawn with replacement per stratum. This definition of pairs is undertaken in a manner closely reflective of the actual design, in that PSUs are paired that are drawn from strata within the same subuniverse, and with similar stratum characteristics. The same definition of pairs was used for each of the assessment components, since all were drawn from the same sample of noncertainty PSUs. The 63 noncertainty PSUs, drawn from 60 strata, were formed into 30 pairs of PSUs, where the pairs were composed of PSUs from adjacent strata within each subuniverse (thus the strata were relatively similar on the characteristics of proportion minority population, population change between 1970 and 1980, and the proportions of urban and farm populations). For those three strata where two PSUs were included in the sample, in each case both PSUs were treated together as constituting a half of one pair. Whereas the actual sample design was to select one PSU with probability proportional to size from each of 60 strata, and then to select supplementary PSUs as needed, for variance estimation purposes the design is regarded as calling for the selection of two PSUs with probability proportional to size with replacement from each of 30 strata. This procedure likely gives a small positive bias to estimates of sampling error.

The student replicate weight for the i^{th} pair of noncertainty PSUs, for the 30 pairs corresponding to values of i from 27 to 56, is computed as follows:

1) Let W_B be the base weight of a student, described in section 10.1.1, which accounts for the various components of the selection probability for the student.

2) At random, one PSU (or set of PSUs from the same stratum) in each pair is denoted as PSU number 1, while the other is denoted as PSU number 2. The i^{th} replicate base weight, W_{Bi}, is given by:

$$W_{Bi} = \begin{cases} 0 & \text{if the student belongs to PSU number 1 of pair i} \\ 2 * W_B & \text{if the student belongs to PSU number 2 of pair i} \\ W_B & \text{if the student is from neither PSU in pair i} \end{cases}$$

3) The i^{th} student replicate weight $SRWT_i$ is obtained by applying the various school and student nonresponse adjustments, the weight trimming, and the poststratification to the i^{th} set of replicate base weights, using procedures identical to those used to obtain the final student weights WT from the set of base weights W_B.

In brief, the procedure for deriving the sets of W_{Bi} value from the W_B values reflects the sampling of PSUs, schools, sessions, and students. By repeating the various weight adjustment procedures in each set of replicate base weights, the impact of these procedures on the sampling variance of the estimator t is appropriately reflected in the variance estimator $\hat{Var}(t)$ defined above.

The procedure for obtaining the 26 sets of replicate weights to estimate the sampling variance from the certainty PSUs is analogous, but somewhat more complex. The first stage of sampling in this case is at the school level, and the derivation of replicate weights must reflect appropriately the sampling of schools within certainty PSUs. Since each of nine different sample components (three age/grade classes by bridge/winter main/spring main) involved different samples of schools, the procedure for forming replicate base weights was individualized to each of these nine sample components. In common across these nine samples were the 34 certainty PSUs used, and the fact that 26 replicate weights were formed in each case.

For a given sample, the 34 certainty PSUs constituted strata, with a sample of schools drawn systematically within each. Using the schools listed in order of sample selection within each stratum, successive schools were paired or formed into triples. These pairs and triples numbered more than 26, so that each replicate weight was in general formed by perturbing the weights of students from more than a single pair or triple. These aggregates of pairs and triples were in general assigned in proportion to the size of the PSU. Thus generally speaking, the four largest PSUs were assigned two replicates each, the next six largest one replicate each, and the remaining 24 were paired and assigned 12 replicates. When splitting the larger PSUs, the schools were split into two groups of (as close as possible) equal size, based on the ordering at the time of sample selection. The first half of the sample was assigned to one replicate, the second half to another. Within each PSU (or half PSU in the case of the four large split PSUs) schools were alternately numbered 1 or 2 starting randomly. If, however, there were exactly three schools sampled in the PSU the schools were randomly numbered 1, 2, or 3. The method of forming replicate base weights in strata where there were not exactly three schools was the same as for the noncertainty strata (except that members of a pair i could come from more than a single "stratum"). When a stratum contained three schools, students in these schools had their weights perturbed for two sets of replicates, say i1 and i2, as follows:

$$W_{Bi1} = \begin{cases} 0 & \text{if the student is in school number 1 of PSU belonging to set i} \\ 1.5 * W_B & \text{if the student is in school number 2 or 3 of a PSU belonging to set i} \\ W_B & \text{if the student does not belong to set i} \end{cases}$$

$$W_{Bi2} = \begin{cases} 1.5 * W_B & \text{if the student is in school number 1 or 2 belonging to set i} \\ 0 & \text{if the student is in school number 3 belonging to set i} \\ W_B & \text{if the student does not belong to set i} \end{cases}$$

The actual pattern of replicate base weight assignment used for each of the nine samples is given in Rust et al. (1992).

The nonresponse, trimming, and poststratification adjustments were applied to each set of replicate base weights to derive the final replicate weights in each case, exactly as in the noncertainty PSUs. In fact these procedures were applied to the full set of weights from all parts of the given sample together, just as for the full sample weights. That is, for example, poststratification factors were derived from the full set of data for each replicate, not separately for certainty and noncertainty PSUs.

This estimation technique was used by NAEP to estimate all sampling errors presented in the various reports. A further discussion of the variance estimation procedure used by NAEP, including a discussion of alternative jackknife estimators that were also considered, appears in Johnson (1989).

We noted above (as discussed in Chapter 11) that a separate estimate of the contribution to variance due to the imprecision in the measure of individual proficiencies is made and added to the jackknife estimate of variance. That variance component could have been appropriately reflected in the jackknife variance estimates simply by separately applying the IRT computations to each jackknife replicate. Because of the heavier IRT computational load, this was not done. Less work was involved by the simple procedure of making separate estimates of this component to be added to the jackknife variance estimates. Also, a separate measure of this component of variance is then available, which would not be so if it were reflected in the jackknife variance estimate.

10.3 APPROXIMATING THE SAMPLING VARIANCE USING DESIGN EFFECTS

In practical terms, the major expenditure of resources in the computation of a jackknife variance estimate occurs in the preparation of estimates for each of the pseudoreplicates. In the 1990 assessment, this implies that the statistic of interest has to be recomputed up to 57 times, once for the overall estimate t, and once for each of the up to 56 pseudoreplicates t_i. Because this is a considerable increase in the amount of computation required, relative to a conventional variance estimate, it is of interest to see how much the jackknife variance estimates differ from their less computationally intensive, simple random sampling based, analogues.

The comparison of the conventional and the jackknife methods of variance estimation will be in terms of a statistic called the *design effect*, which was developed by Kish (1965) and extended by Kish and Frankel (1974). The design effect for a statistic is the ratio of the actual variance of the statistic (taking the sample design into account) over the conventional variance estimate based on a simple random sample with the same number of elements. The design effect is the inflation factor to be applied to the conventional variance estimate in order to adjust error estimates based on simple random sampling assumptions to account approximately for the effect of the sample design. The value of the design effect depends on the type of statistic computed and the variables considered in a particular analysis as well as the combined clustering, stratification, and weighting effects occurring among sampled elements. Generally, the design effects for statistics from complex samples such as NAEP are greater than one, because variances based on simple random sampling assumptions tend to provide underestimates of the variances of statistics calculated from complex samples.

10.3.1 Design Effects for Proportion-correct Statistics

As an example of the distribution of design effects to be expected from NAEP data, we consider the design effect for the statistic P, the estimated proportion of a specified subgroup of the population who would correctly respond to a given assessment item. The proportion-correct statistic is the weighted mean of the responses to the item of the assessed individuals who belong to the subgroup, where an individual's response is either 1=correct or 0=incorrect. The design effect for the proportion-correct statistic P is of the form

$$\text{deff}(P) = [\text{Var}_{JK}(P)]/[P(1 - P)/N].$$

In the above, N is the total number of individuals in the subgroup responding to the item, $\text{Var}_{JK}(P)$ is the jackknife variance of P, and $P(1 - P)/N$ is the conventional variance estimate of P. (Although the estimate $P(1 - P)/N$ has the same form as the simple random sampling estimator of the variance of a proportion correct, the use of sample weights in the estimation of P reflects the appropriate distribution of the population.)

The design effects for the proportion-correct statistics for each item administered in a sample are summarized by the mean, median, lower quartile, upper quartile, and standard deviation across proportion-correct statistics for each item in the sample. These summaries of the design effects are given for each sample in Appendix D. The tables in Appendix D also contain degree of freedom estimates explained in section 10.4. The numerous tables in Appendix D have been further summarized in Tables 10-6 through 10-11 by averaging the entries in Appendix D across all main samples to produce one summary table for each age/grade level and likewise for all bridge and special mathematics samples. Table 10-12 summarizes the design effects for the 1990 Trial State Assessment of mathematics for comparison with the national sample results. The particular demographic variables shown were selected because (1) they are major variables in NAEP reports and (2) they reflect different types of divisions of the population that might have different levels of sampling variability.

There are some systematic differences in the design effects for different types of samples and different subpopulations. The bridge samples tend to have larger design effects, especially those administered in tape sessions, because the same items must be administered to each student in a tape session so that the number of students per school administered an item is higher for samples with tape sessions than for samples of similar size with spiraled administrations. This results in more clustering since students within a school tend to have more similar performance than students chosen in a simple random sample. Samples for the state assessment tend to have smaller design effects than the national samples. However, the properties of the state samples vary considerably.

The estimates for the total population tend to have the largest design effects, while the estimates for subpopulations, such as those based on parents' education level, tend to have smaller design effects. The parent education categorization forms more homogeneous collections of schools that have students with more similar backgrounds and performance thus reducing the variation in the types of schools (and students) included in the population estimates. The samples assessing reading items also tend to have smaller design effects. One conjecture that could account for this observation is that reading items may be less curriculum dependent and are thus less dependent on specific course offerings that vary among schools.

Table 10-6

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics Averaged Across Main Samples
Grade 4

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.23	1.47	1.75	2.72	1.51	0.35	38
Male	1.19	1.41	1.62	2.49	1.44	0.34	36
Female	1.22	1.42	1.63	2.40	1.43	0.34	36
White	1.12	1.37	1.62	2.49	1.40	0.35	34
Black	1.02	1.18	1.45	2.47	1.24	0.33	28
Hispanic	1.11	1.30	1.52	2.47	1.33	0.31	37
Asian American	0.89	1.12	1.72	5.88	1.48	0.96	5
Other Race/Eth.	1.00	1.25	1.49	2.50	1.28	0.37	24
Other Metro	1.22	1.47	1.76	2.94	1.52	0.39	31
Low Metro	1.03	1.24	1.51	2.52	1.30	0.39	23
High Metro	1.09	1.42	1.73	3.26	1.48	0.53	16
Par. Ed. < HS	1.04	1.23	1.56	3.00	1.33	0.42	19
Par. Ed. = HS	1.09	1.30	1.60	2.96	1.38	0.43	22
Par. Ed. > HS	1.10	1.27	1.52	2.42	1.33	0.32	35
Par. Ed. = College	1.19	1.38	1.57	2.82	1.40	0.33	37
Par. Ed. = IDK	1.15	1.35	1.58	2.30	1.38	0.32	37
Public School	1.19	1.44	1.68	2.70	1.45	0.35	36
Nonpublic School	1.22	1.42	1.70	2.63	1.47	0.36	34

Table 10-7

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics Averaged Across Main Samples
Grade 8

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.29	1.52	1.83	3.40	1.59	0.44	27
Male	1.20	1.44	1.69	2.56	1.45	0.35	35
Female	1.20	1.40	1.67	2.56	1.44	0.36	33
White	1.19	1.45	1.74	3.39	1.50	0.44	25
Black	1.06	1.25	1.52	2.46	1.31	0.35	28
Hispanic	1.09	1.31	1.59	2.42	1.37	0.38	26
Asian American	1.03	1.31	1.67	4.14	1.43	0.58	12
Other Race/Eth.	0.87	1.15	1.79	5.16	1.46	0.87	6
Other Metro	1.26	1.47	1.71	3.02	1.52	0.41	28
Low Metro	1.07	1.49	1.85	4.20	1.57	0.67	11
High Metro	1.00	1.55	2.65	9.14	2.06	1.57	3
Par. Ed. < HS	1.00	1.21	1.46	2.26	1.24	0.34	28
Par. Ed. = HS	1.02	1.21	1.46	2.64	1.26	0.34	26
Par. Ed. > HS	1.06	1.20	1.40	2.39	1.24	0.30	34
Par. Ed. = College	1.18	1.41	1.71	2.61	1.46	0.39	29
Par. Ed. = IDK	1.10	1.29	1.53	2.37	1.32	0.32	33
Public School	1.20	1.46	1.77	3.36	1.52	0.45	25
Nonpublic School	1.31	1.54	1.89	3.74	1.64	0.53	20

Table 10-8

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics Averaged Across Main Samples
Grade 12

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.34	1.57	1.87	3.14	1.62	0.42	30
Male	1.21	1.49	1.73	2.74	1.50	0.39	30
Female	1.17	1.36	1.61	2.36	1.40	0.32	38
White	1.26	1.52	1.83	3.26	1.58	0.45	26
Black	1.02	1.24	1.48	2.45	1.27	0.34	28
Hispanic	1.09	1.38	1.68	2.96	1.43	0.44	22
Asian American	0.96	1.24	1.95	6.95	1.61	1.07	5
Other Race/Eth.	0.90	1.16	1.52	3.50	1.24	0.48	14
Other Metro	1.27	1.53	1.87	3.01	1.59	0.43	28
Low Metro	1.17	1.68	2.58	6.82	1.98	1.18	6
High Metro	1.09	1.45	1.94	3.66	1.56	0.63	12
Par. Ed. < HS	0.99	1.20	1.46	2.39	1.25	0.35	27
Par. Ed. = HS	1.07	1.31	1.57	2.42	1.35	0.37	26
Par. Ed. > HS	1.12	1.32	1.53	2.48	1.35	0.33	33
Par. Ed. = College	1.17	1.36	1.62	2.42	1.40	0.33	36
Par. Ed. = IDK	1.00	1.25	1.49	3.09	1.30	0.41	22
Public School	1.21	1.44	1.74	2.71	1.49	0.38	31
Nonpublic School	1.37	1.70	2.04	3.29	1.75	0.52	23

Table 10-9

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics Averaged Across Trend Samples
Age 9

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.52	1.80	2.10	3.57	1.86	0.52	28
Male	1.33	1.53	1.82	2.56	1.58	0.39	33
Female	1.28	1.53	1.84	3.02	1.59	0.42	32
White	1.26	1.48	1.76	3.18	1.54	0.44	29
Black	1.18	1.48	1.91	3.75	1.58	0.56	17
Hispanic	1.13	1.36	1.65	3.14	1.42	0.45	20
Asian American	0.93	1.35	2.14	5.98	1.66	1.06	6
Other Race/Eth.	0.91	1.08	1.29	2.36	1.13	0.33	28
Other Metro	1.30	1.73	2.07	3.52	1.78	0.53	27
Low Metro	1.41	1.96	2.82	7.48	2.31	1.36	8
High Metro	1.00	1.39	1.94	4.49	1.58	0.82	9
Par. Ed. < HS	0.96	1.15	1.36	2.17	1.18	0.31	30
Par. Ed. = HS	1.03	1.30	1.55	2.87	1.34	0.42	23
Par. Ed. > HS	1.00	1.21	1.41	2.27	1.24	0.32	31
Par. Ed. = College	1.29	1.52	1.79	2.85	1.56	0.41	32
Par. Ed. = IDK	1.17	1.36	1.60	2.50	1.40	0.34	33
Public School	1.51	1.77	2.10	3.50	1.84	0.52	28
Nonpublic School	1.25	1.68	2.31	6.03	1.93	1.02	8

Table 10-10

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics Averaged Across Trend Samples
Age 13

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.37	1.68	2.16	3.39	1.78	0.57	24
Male	1.19	1.44	1.80	3.03	1.51	0.46	23
Female	1.21	1.46	1.76	2.86	1.51	0.41	28
White	1.21	1.46	1.84	3.35	1.56	0.53	21
Black	1.04	1.36	1.80	3.97	1.49	0.61	15
Hispanic	1.05	1.30	1.63	3.15	1.37	0.47	19
Asian American	0.98	1.51	2.43	8.65	1.98	1.50	4
Other Race/Eth.	0.84	1.08	1.50	3.59	1.22	0.52	16
Other Metro	1.34	1.64	2.08	3.49	1.73	0.58	22
Low Metro	1.07	1.59	2.12	6.24	1.78	1.06	7
High Metro	0.99	1.47	2.12	5.13	1.67	0.93	7
Par. Ed. < HS	0.98	1.19	1.48	2.78	1.26	0.39	22
Par. Ed. = HS	1.06	1.27	1.50	2.35	1.29	0.32	34
Par. Ed. > HS	1.05	1.24	1.46	2.38	1.28	0.32	35
Par. Ed. = College	1.18	1.49	1.76	3.23	1.53	0.47	24
Par. Ed. = IDK	1.03	1.21	1.47	2.28	1.26	0.33	29
Public School	1.34	1.69	2.15	3.54	1.78	0.60	24
Nonpublic School	0.99	1.36	1.96	3.70	1.51	0.72	12

Table 10-11

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics Averaged Across Trend Samples
Age 17

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.36	1.63	1.97	3.97	1.70	0.49	28
Male	1.24	1.47	1.77	3.17	1.53	0.42	29
Female	1.20	1.44	1.73	3.00	1.50	0.42	29
White	1.18	1.43	1.76	4.17	1.53	0.54	21
Black	1.14	1.49	1.91	4.53	1.61	0.66	18
Hispanic	0.98	1.27	1.72	3.80	1.40	0.60	12
Asian American	0.94	1.26	1.70	4.29	1.40	0.68	9
Other Race/Eth.	0.90	1.22	1.55	3.10	1.26	0.49	14
Other Metro	1.24	1.48	1.85	3.71	1.57	0.49	23
Low Metro	1.30	2.00	2.91	7.17	2.27	1.32	8
High Metro	1.23	1.73	2.50	8.05	2.05	1.24	9
Par. Ed. < HS	1.02	1.26	1.51	3.40	1.31	0.43	19
Par. Ed. = HS	1.08	1.29	1.55	2.88	1.35	0.38	25
Par. Ed. > HS	1.11	1.32	1.60	2.95	1.38	0.40	24
Par. Ed. = College	1.16	1.39	1.67	3.45	1.45	0.44	25
Par. Ed. = IDK	1.02	1.25	1.51	3.00	1.32	0.46	18
Public School	1.29	1.56	1.90	3.68	1.63	0.48	26
Nonpublic School	1.23	1.79	3.28	10.12	2.48	1.84	8

Table 10-12

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics Averaged Across State Samples
Grade 8

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.08	1.27	1.49	2.55	1.31	0.34	32
Male	1.02	1.20	1.40	2.22	1.23	0.28	40
Female	1.03	1.20	1.40	2.34	1.23	0.30	37
White	1.03	1.22	1.44	2.59	1.27	0.35	30
Black	0.84	1.08	1.37	3.26	1.21	0.57	19
Hispanic	0.93	1.11	1.32	2.23	1.14	0.31	30
Asian American	0.89	1.07	1.29	2.43	1.11	0.35	25
Other Race/Eth.	0.92	1.11	1.33	2.56	1.15	0.36	28
Other Metro	1.06	1.27	1.51	2.63	1.31	0.36	28
Low Metro	0.72	1.07	1.50	4.02	1.20	0.68	8
High Metro	0.72	1.05	1.53	3.97	1.20	0.68	8
Par. Ed. < HS	0.94	1.10	1.28	2.11	1.13	0.28	35
Par. Ed. = HS	0.99	1.15	1.34	2.03	1.17	0.26	41
Par. Ed. > HS	0.97	1.12	1.29	2.04	1.14	0.25	43
Par. Ed. = College	1.01	1.17	1.37	2.16	1.20	0.28	38
Par. Ed. = IDK	0.95	1.11	1.28	2.10	1.12	0.26	38

The tables show that the design effects are predominantly larger than 1, indicating that standard variance estimation formulas will be generally too small, sometimes markedly so. Although the distributions of design effects appear somewhat different for certain subgroups of the population, they are, perhaps, similar enough (at least within a grade) to select an overall composite value that is adequate for most purposes. In choosing a composite design effect, some consideration must be given to the relative consequences of overestimating the variance as opposed to underestimating the variance. For example, adopting the position that an overestimate of the variance is as severe an error as an underestimate leads to using a composite that is near to the center of the distributions of the design effects. Possible composites of this type are the mean and median design effects across the combined distribution of all design effects. In the current data, the mean design effects for total population estimates (which tend to be larger than most subpopulation estimates) are 1.51, 1.59, and 1.62 for main samples and 1.86, 1.78, and 1.70 for bridge and special mathematics samples respectively for grades 4, 8, and 12. These are close to, but greater than, the median design effects: 1.47, 1.52, and 1.57 for main samples and 1.80, 1.68 and 1.63 for bridge and special mathematics samples.

Alternatively, one can adopt the position that it is a graver error to underestimate the variability of a statistic than to overestimate it. For example, Johnson and King (1987) examine estimation of variances using design effects (among other techniques) under the assumption that the consequences of an underestimate are three times as severe as those of an overestimate of the same magnitude. Adopting a loss function that is a weighted sum of absolute values of the deviations of predicted from actual, with underestimates receiving three times the weight of overestimates, produces the upper quartile of the design effects as the composite value. This assumes that the size of the design effects do not depend on the size of the variance estimates. The values of this composite, for grades 4, 8, and 12, respectively, are 1.75, 1.83, and 1.87 for the main samples and 2.10, 2.16, and 1.97 for the bridge and special mathematics samples.

10.4 THE DEGREES OF FREEDOM OF THE VARIANCE ESTIMATE

It is important to have an indication of the number of degrees of freedom to attribute to the jackknife variance estimator $\hat{Var}(t)$. The degrees of freedom of a variance estimator provide information on the stability of that estimator: the higher the number of degrees of freedom, the lower the variability of the estimator. In practical terms, the number of degrees of freedom of the variance estimator corresponds to the number of residual degrees of freedom that can be assumed for inferential procedures.

Since the jackknife procedure estimates the sampling variability of the statistic by assessing the effect of change in the sample at the paired first-stage sampling unit (FSSU) level, the number of degrees of freedom of the variance estimator $\hat{Var}(t)$ will be at most equal to M, the number of FSSU pairs. The maximum number of degrees of freedom equals the number of independent pieces of information used to generate the variance. In the case of data from the main assessments, the pieces of information are the 56 squared differences $(t_i - t)^2$, each supplying at most one degree of freedom (regardless of how many individuals were sampled within any FSSU).

The number of degrees of freedom of the sample variance estimator can be strictly less than the number of FSSU pairs. For example, suppose that the statistic t is a mean for some

subgroup and no members of that subgroup can come from either FSSU in the i^{th} FSSU pair. (Examples of such a subgroup are any PSU-level partitioning of the population, such as region.) In this instance, neither member of the FSSU pair i directly contributes to the estimate of t, so that the pseudoreplicate t_i would nearly equal the statistic t. If the replicate weights used to generate t_i had not received poststratification adjustments, the resulting pseudoreplicate t_i would be identical to the overall estimate t so that $(t_i - t)^2 = 0$. In this case, such a FSSU pair would impart no information to the variability of the statistic t and thus contribute zero degrees of freedom to the variance. Howeve-, since the replicate weights have received poststratification adjustments, the component $(t_i - t)^2$ is measuring the effect of the poststratification on the estimate. While being nonzero, such a component will tend to be much smaller in magnitude than the squared difference $(t_k - t)^2$ for any PSU pair k that does contribute to the estimate of t (see Rust, 1985).

The squared difference $(t_i - t)^2$ estimates σ_i^2, say, the contribution to the sampling variance of the statistic t which can be attributed to the i^{th} FSSU pair and $\hat{Var}(t)$ estimates the sum of the contributions across all pairs:

$$\sum_{i=1}^{M} \sigma_i^2 \ .$$

If the σ_i^2 vary widely, as when a few of the σ_i^2 are markedly larger than the remainder, as in the above case where neither member of an FSSU pair contributes to the estimate of t, then $\hat{Var}(t)$ is predominantly estimating the sum of these larger components, which dominate the remaining terms. The effective degrees of freedom of $\hat{Var}(t)$ in this case will be nearer to the number of dominant terms.

One way to estimate how many degrees of freedom to attribute to the jackknife variance estimate of a statistic t is to match estimates of the first two moments of $\hat{Var}(t)$ to those of a chi-square random variable (Satterthwaite, 1941). If the t_i are normally distributed, the effective number of degrees of freedom using this approximation is

$$df_{eff} = \frac{\left(\sum_{i=1}^{M} (t_i - t)^2\right)^2}{\sum_{i=1}^{M} (t_i - t)^4}.$$

However, empirical evidence from simulations indicates that the above formula has a severe downward bias in the case of the sum of single degree of freedom chi-square random variables (Johnson & Rust, in press). More direct ways of assessing the effective degrees of freedom of a variance estimate are possible when a number of independent replicates of the estimate are available.

It is possible to estimate the number of degrees of freedom to attribute to the jackknife variance estimates of the weighted proportion-correct statistics by considering the distribution of design effects for a given set of items in a population or subpopulations (such as males or total) under the assumptions that the individual design effects are all estimating the same, underlying,

design effect D and that the variance estimates of all weighted proportion-correct statistics have the same degrees of freedom, f. Specifically, assume that the jackknife variance estimate, V_j, of the j^{th} weighted proportion-correct statistic, P_j, is distributed like the random variable $(\sigma_j^2 / f) X_f^2$, where X_f^2 is a chi-square random variable with f degrees of freedom and σ_j^2 is the expected value of V_j. Further assume that the expected value of the conventional variance estimate is σ_j^2/D, where D is the underlying design effect. Then, for a sufficiently large sample size, so that the conventional variance estimate can be taken to be σ_j^2/D, the design effect of P_j will be approximately distributed like the constant (D/f) times a chi-square random variable with f degrees of freedom. If the underlying design effect D and the degrees of freedom f are the same for all P_j, then the distribution of the estimated design effects of the proportions correct across the set of items will be approximately distributed like a multiple times a chi-square random variable with f effective degrees of freedom.

From here, Johnson (1989) proceeds by assuming that the design effects $D_1, \ldots, D_I$ across I items are independent estimates of a common design effect, D, and forms quantile plots of D_j against a χ_f^2 distribution, for f = 1, . . ., M degrees of freedom, where M is the number of jackknife pseudoreplicates. Regression lines are fit through the origin for each choice of f, and the value of f with the best fitting line (as measured by residual mean square) is used as the estimate of f the effective degrees of freedom. The quality of this approximation to the distribution of design effects is quite good. In every case examined by Johnson (1989), the prediction of the design effects with a chi-square distribution with the effective degrees of freedom accounted for at least 94 percent of the variance of the actual values.

A computationally simpler estimate can also be obtained from the same approximation by matching moments with the empirical distribution of $D_1, \ldots, D_I$,

$$\bar{D} \approx E\left(\frac{D}{f}\chi_f^2\right) = \frac{D}{f} f = D$$

$$\hat{\sigma}_D^2 \approx \text{var}\left(\frac{D}{f}\chi_f^2\right) = \frac{D^2}{f^2}(2f) = \frac{2}{f}D^2$$

which implies

$$\hat{\sigma}_D^2 = 2\frac{(\bar{D})^2}{f}$$

or

$$\hat{f} = 2\left(\frac{\bar{D}}{\hat{\sigma}_D}\right)^2 .$$

If the D_j are positively correlated, $\hat{f}$ will tend to overestimate f because $\hat{\sigma}_D^2$ underestimates the variance of D_j. To see this, let

$$\mu_D = E(D_j),$$
$$\sigma_D = \sqrt{\mathrm{var}(D_j)},$$

which are assumed to be the same for $j = 1, \ldots, I$, and center D_j by μ_D,

$$d_j = D_j - \mu_D.$$

Then

$$E(\hat{\sigma}_d^2) = \frac{1}{I-1} \sum_{i=1}^{I} \left[E(d_i)^2 - 2E(d_i \bar{d}) + E(\bar{d})^2 \right]$$

$$= \frac{1}{I-1} \left\{ \sum_{i=1}^{I} \left(E(d_i^2) - \frac{2}{I} E(d_i^2) + \frac{1}{I} E(d_i^2) \right) \right.$$

$$\left. - \frac{2}{I} \sum_{i=1}^{I} \sum_{\substack{j=1 \\ j \neq I}}^{I} E(d_i\, d_j) + \frac{2}{I} \sum_{i=1}^{I-1} \sum_{j=i+1}^{I} E(d_i\, d_j) \right\}$$

$$= \sigma_D^2 + \frac{1}{I-1} \left\{ \frac{-4}{I} \sum_{i=1}^{I-1} \sum_{j=i+1}^{I} E(d_i\, d_j) + \frac{2}{I} \sum_{i=1}^{I-1} \sum_{j=i+1}^{I} E(d_i\, d_j) \right\}$$

$$= \sigma_D^2 - \frac{2}{(I-1)I} \sum_{i=1}^{I-1} \sum_{j=i+1}^{I} \mathrm{Cov}(D_i, D_j),$$

so that if all or most of the D_i are positively correlated as we anticipate, our estimate of σ_D^2 will be too small, while D is still estimating μ_D. Estimates of f based on quantiles also appear to have this undesirable feature, but it is difficult to obtain a definitive answer because of the complexity of these estimates.

This estimator is in good agreement with the quantile-based estimate in Johnson (1989). The degree of freedom estimates for grade 12 in Table 8-11 of Johnson, Rust, and Hansen (1990, p. 224) were replicated using the summary statistics for design effects given in Table 8-8 (p. 216) of that report. The only complication was that $\hat{\sigma}$ was not included in Table 8-8, so it was approximated using 3/4 (HiQ - LoQ). Since the D_i have a skewed distribution, this approximation likely understates $\hat{\sigma}_D$ and thus our reported estimates are larger than our $\hat{f}$ would be if computed with $\hat{\sigma}_D$. The results using the moment matching and the quantile-based estimates from Table 8-11 of Johnson, Rust, and Hansen (1990) are given in Table 10-13.

The tables in Appendix D show the result of this estimation of the effective degrees of freedom of the design effects, and hence the jackknife variance estimates, for weighted proportion-correct statistics for all samples in the 1990 assessment. The effective degrees of freedom in these tables are summarized in Tables 10-6 through 10-12.

Table 10-13

Effective Degrees of Freedom for the Design Effects of the Proportion-correct Statistics

Group	Quantile-based	Matching Moments
Total	40	47
Male	35	41
Female	40	42
White	41	43
Black	30	31
Hispanic	30	38
Other	33	29
< Modal Grade	13	10
At Modal Grade	41	32
> Modal Grade	40	43
Northeast	11	18
Southeast	9	12
Central	10	11
West	13	16
Rural	7	7
Low Metro	13	17
High Metro	13	11
Big City	13	17
Fringe	11	9
Medium City	9	8
Small Place	18	14
< High School	32	31
= High School	36	49
High School +	43	51
Grad. College	44	35
IDK	14	20
Public School	42	51
Nonpublic School	14	14

The numbers in the tables show that the effective degrees of freedom of the jackknife variance estimates are indeed no larger than the number of FSSU pairs, and are, in fact, markedly smaller in some cases. The Asian American population is a example of a subgroup that has consistently small estimates of degrees of freedom. This is due to the fact that this population is concentrated in relatively few of the primary sampling units.

The effective degrees of freedom for the NAEP jackknife variance estimates are much smaller than the degrees of freedom attributed to the corresponding error estimates from conventional techniques. This fact affects inferential procedures since significance tests based on the conventional degrees of freedom will be too liberal (and confidence intervals will be too small). Fortunately, for the usual significance levels, the effect of using the effective degrees of freedom rather than the conventional values is generally moderate: a t statistic significant at the $\alpha=5\%$ level assuming infinite degrees of freedom (essentially the conventional estimate) is significant at the $\alpha=6\%$ level for 20 effective degrees of freedom, the $\alpha=7\%$ level for 10 effective degrees of freedom, and the $\alpha=10\%$ level for five effective degrees of freedom.

For practical purposes, the impact of the reduced degrees of freedom on inferential techniques can be largely accounted for by (1) using a moderate number (say 25) of degrees of freedom for all inferences about subgroups that appear approximately uniformly in all PSUs, and (2) using a smaller number (say 10) for the remaining subgroups. Certainly one should be cautious about barely significant results for subgroups that are highly clustered in the population.

Chapter 11

SCALING PROCEDURES[1]

Robert J. Mislevy

Educational Testing Service

The primary method by which results from the 1990 National Assessment of Educational Progress were disseminated is scale-score reporting. With scaling methods, the performance of a sample of students in a subject area or subarea can be summarized on a single scale or series of subscales even when different students have been administered different items. This chapter presents an overview of the scaling methodologies employed in the analyses of the data from NAEP surveys:

- Section 11.1 briefly discusses the perspective on scaling from which the procedures were conceived and applied.
- Section 11.2 reviews the "plausible values" methodology used in NAEP scale-score analyses.
- Section 11.3 describes how plausible values are used in subsequent analyses.
- Section 11.4 lists the scale-score analyses carried out on the 1990 data.

Details of the scaling procedures specific to the subject areas of reading, mathematics, and science are presented in Chapters 12, 13, and 14.

11.1 BACKGROUND

NAEP reports were originally envisaged some 20 years ago as simple lists of percents correct to individual survey items, in the population as a whole and in subpopulations of particular interest. It soon became apparent that major features of the detailed results from hundreds of items could not be effectively communicated without some kind of summarization. Averaging percents-correct from individual items summarizes results, but limits comparisons to groups of items that are common over the time points or student subpopulations that are to be compared. These limitations can be overcome by the use of response scaling methods. If

[1] The contributions of Nancy Allen, Albert Beaton, James Carlson, David Freund, Eugene Johnson, Bruce Kaplan, Jennifer Nelson, Kathleen Sheehan, Neal Thomas, Minhwei Wang, Kentaro Yamamoto, and Rebecca Zwick are gratefully acknowledged.

several items require similar skills, the regularities observed in response patterns can often be exploited to characterize both respondents and items in terms of a relatively small number of variables. When combined through the scaling model, these variables capture the dominant features of the data. Using the scale, it becomes possible to talk about distributions of proficiency in a population or subpopulation, and to estimate the relationships between proficiency and background variables. Item response theory (IRT; see Hambleton, 1989, for an overview) and a newly developed procedure called the average response method (ARM), both of which are reviewed in section 11.2, are the two scaling procedures ETS has employed in NAEP reporting to date.

Of course, any procedure of aggregation, from a simple average to a complex multidimensional scaling model, highlights certain patterns at the expense of other potentially interesting patterns that may reside within the data. In a very real sense, every single item in a NAEP survey is of interest in its own right, and can provide useful information about what young Americans know and can do. The choice of an aggregation procedure must be driven by a conception of just which patterns are salient for a particular purpose. The procedure that is optimal for one purpose may be poorly suited for another. The relatively high levels of aggregation found in ETS/NAEP reports such as *The Reading Report Card: Progress Toward Excellence in Our Schools* (NAEP, 1985a), for example, are well suited to general discussions of trends and policy implications. These reports average over, and are therefore *not* keyed to, the microanalysis of performance at the level of specific skills; neither do they reveal popular student misconceptions or erroneous rules, as might be of interest to classroom teachers in a subject area. By no means do the scale-score methods we employ as a reporting vehicle exhaust the potential of NAEP data. Indeed, NAEP's secondary-use data files, which contain the original responses of all surveyed students, were created expressly to encourage secondary researchers to carry out alternative analyses from different perspectives.

The reporting scale of a NAEP survey, then, simply summarizes performance on a collection of educational tasks in much the same way that the Consumer Price Index (CPI) summarizes the total cost of a market basket of products. The two indices exhibit some of the same useful features and limitations. Just as the CPI composite represents average American spending patterns, the items in a NAEP survey were specified by independent consensual process to tap a "market basket of skills." Just as changes in the CPI reflect at a glance changes in the cost of goods in general, changes in NAEP scale-score distributions reflect changes in proficiency as averaged over the items in the pool. But understanding just how and why the CPI changes requires deeper analyses, into specific components of the market basket; when the CPI goes up, some of the components will have gone up by greater rates than others, while some may have even dropped in price. The NAEP scale depends similarly on the balance of items of varying types and topics in the survey, and reflects only an average over the varying patterns among them.

NAEP first attempts to carry out scaling in subject areas in which similar patterns can be expected over items; for example, within five more narrowly defined topics within mathematics, such as numbers and operations, measurement, and algebra and functions. Carrying out scaling in separate subareas captures trends and comparisons that may differ across subareas—for example gender differences in mathematics performance by subscale—for reasons that can include different curricular emphases over time or across schools. As is done in the 1990 NAEP surveys of mathematics and science, these subscale results are supplemented by a subject area

average, or "composite." A composite is computed as the weighted average of the subscale scores where the weights correspond to the relative importance given to each subscale as defined by the objectives. This is comparable to calculating price changes in separate market baskets for food, transportation, energy, and so on, and reporting these individually along with the overall average. Then, within each scaling area, NAEP highlights meaningful departures from general trends in several ways, such as (1) supplementing scale-score distributional results with more detailed breakdowns in terms of percents correct for groups of related items, and (2) explicating countertrends or comparisons that can be identified with one or a few items, or with particular subpopulations of students. This is analogous to reporting that the Consumer Price Index jumped 5 percent, but noting that the increase was mainly due to a change in oil prices.

The basic information from an assessment consists of the responses of students to the items presented in the assessment. For NAEP, these items are generated to measure performance on sets of objectives developed by nationally representative panels of learning area specialists, educators, and concerned citizens. Satisfying the objectives of the assessment and ensuring that the tasks selected to measure each goal cover a range of difficulty levels typically requires a large number of items. To reduce student burden, each assessed student was presented only a fraction of the full pool of items using multiple matrix sampling procedures.

11.2 SCALING METHODOLOGY

This section reviews the scaling models employed in the analyses of 1990 NAEP data, as well as the "plausible values" methodology that allows such models to be used with NAEP's sparse item-sampling design. The reader is referred to Mislevy (1991) for an introduction to plausible values methods and a comparison with standard psychometric analyses, to Mislevy and Sheehan (1987), Beaton and Johnson (1990), and Mislevy, Johnson, and Muraki (in press) for additional information on how the models are used in NAEP, and to Rubin (1987) for the theoretical underpinnings of the approach.

11.2.1 The Scaling Models

Two types of scaling models have been used by NAEP in recent assessments. The three-parameter logistic (3PL) model from item response theory (IRT; see Lord, 1980) was used in 1990 for the subject areas of reading, mathematics, and science. The average response method (ARM; Beaton & Johnson, 1987, 1990), an extension of multiple regression developed by NAEP for the 1984 assessment, was used for the subject area of writing and for summarizing background information and attitude responses in assessments prior to 1990. The 3PL and the ARM are both "latent variable" models, quantifying respondents' tendencies to provide responses in a given direction (e.g., correct answers to items in a subject area; positive responses on attitude questions; higher rather than lower ratings in written essays), as a function of a parameter that is not directly observed.

The three-parameter logistic (3PL) IRT model. The fundamental equation of the 3PL model is the probability that a person whose proficiency on subscale k is characterized by the *unobservable* variable θ_k will respond correctly to item j:

$$P(x_j = 1 \mid \theta_k, a_j, b_j, c_j) = c_j + (1 - c_j)/\{1 + \exp[-1.7 a_j(\theta_k - b_j)]\}$$

$$\equiv P_j(\theta_k), \tag{11.1}$$

where

x_j is the response to item j, 1 if correct and 0 if not;

a_j where $a_j > 0$, is the slope parameter of item j, characterizing its sensitivity to proficiency;

b_j is the threshold parameter of item j, characterizing its difficulty; and

c_j where $0 \leq c_j < 1$, is the lower asymptote parameter of item j, reflecting the chances of a correct response from students of very low proficiency; c parameters are estimated for multiple-choice items, but are fixed at zero for open-ended items.

For the purposes of reporting item parameter estimates and other intermediary estimates, the linear indeterminacy apparent in (11.1) may be resolved by an arbitrary choice of the origin and unit size in a given scale. For example, a provisional scale was employed in the analysis of the 1984 reading assessment by standardizing the combined age 9/grade 4, age 13/grade 8, and age 17/grade 11 samples. To aid interpretation, final published results were linearly transformed from the θ scale to a 0-to-500 "reading proficiency scale" (Beaton, 1987). Analogous scaling conventions and reporting transformations for the 1990 assessment are described in the corresponding subject area chapters in this report.

A typical assumption in IRT is the conditional independence of the probabilities of correct response by an individual to a set of items, given the individual's proficiency. That is, conditional on the individual's θ_k, the joint probability of a particular response pattern $\underline{x} = (x_1, \ldots, x_n)$ across a set of n items is simply the product of terms based on (11.1):

$$P(\underline{x} \mid \theta_k, \underline{a}, \underline{b}, \underline{c}) = \prod_j^n [P_j(\theta_k)]^{x_j} [1 - P_j(\theta_k)]^{1 - x_j} \tag{11.2}$$

Furthermore, it is also typically assumed that response probabilities are conditionally independent of background variables ($\underline{y}$), given θ_k, or

$$P(\underline{x} \mid \theta_k, \underline{a}, \underline{b}, \underline{c}, y) = p(\underline{x} \mid \theta_k, \underline{a}, \underline{b}, \underline{c}).$$

After $\underline{x}$ has been observed, equation 11.2 can be viewed as a likelihood function, and provides a basis for inference about θ_k or about item parameters. Estimates of item parameters were obtained with a modified version (Rogers & Nelson, 1990) of Mislevy and

Bock's (1982) BILOG computer program, then treated as known in subsequent calculations. In subject areas with multiple subscales, the parameters of the items constituting each subscale were estimated independently of the parameters of the other subscales. Once items have been calibrated, a likelihood function for the subscale proficiency θ_k is induced by a vector of responses to any subset of calibrated items, thus allowing θ-based inferences from matrix samples.

Conditional independence is a mathematical assumption, not a necessary fact of nature. Although the IRT models are employed in NAEP only to summarize average performance, a number of checks are made to detect serious violations of conditional independence, and, when warranted, remedial efforts are made to mitigate its effects on inferences. These checks include the following:

1) Checks on relative item operating characteristics among distinct gender and ethnicity groups (i.e., differential item functioning, or DIF). Some degree of relative differences are to be expected, of course, and modestly varying profiles among groups will exist beyond the differences conveyed by their differing θ distributions. The intent of the check at this stage is to detect and eliminate items that operate differentially for identifiable reasons that are unrelated to the skills intended to be measured in the subject area.

2) When a subscale extends over age groups, as is the case for the national mathematics subscales, evidence is sought of different operating characteristics over ages. When such effects are found, an item in question is represented by different item parameters in different age groups. For such an item, the probability of a correct response given θ depends on the age group in question—a departure from conditional independence incorporated into the model in the interest of fidelity to the data. This is analogous to calibrating items separately in different age groups, and linking the resulting scales via those items whose response curves in the separate ages can be rectified by a single linear transformation.

Item-level factor analyses have diminished in importance as our perspective of the role of IRT in NAEP has evolved. The assumption that performance in a scaling area is driven by a single unidimensional variable is unarguably incorrect in detail. However, our use of the model is not theoretical, instead it is data analytic; interpretation of results is not trait-referenced, but domain-referenced. Scaling areas are determined *a priori* by considerations of content as collections of items for which overall performance is deemed to be of interest. The IRT summary is not expected to capture all meaningful variation in item response data, but to reflect distributions of overall proficiency—to summarize the main patterns in item percents-correct in the populations and subpopulations of interest. Using a unidimensional IRT model when the true model is multidimensional captures these overall patterns even though it over- or under-estimates the covariances among pairs of items. For inferences based on overall proficiency, violations of the model with respect to dimensionality are less serious than violations in the shapes of the marginal response curves enumerated above—hence our greater attention to routine checks of item-fit residuals for every item in every calibration run than to factor analytic results.

The local independence assumption embodied in (11.2) implies that item response probabilities depend only on θ and the specified item parameters—not on the position of the item in the booklet, on the content of items around an item of interest, or on test-administration timing conditions. These effects are certainly present in any application. The practical question is whether the IRT probabilities obtained via (11.2) are "close enough" to be robust with respect to the context in which the data are to be collected and the inferences that are to be drawn.

Experience with adaptive testing has shown using the same item parameters regardless of when an item is administered does not materially bias estimates of the proficiencies of individual examinees. Our experience with the 1986 NAEP reading anomaly, has shown, however, that for measuring small changes over time, changes in item context and speededness conditions lead to unacceptably large random error components (Beaton & Zwick, 1990). These can be avoided by presenting items used to measure change in identical test forms, with identical timings and administration conditions. Thus we do *not* maintain that the item parameter estimates obtained in any particular booklet configuration are appropriate for other conceivable configurations, and acknowledge that the parameter estimates are context-bound. For this reason, we prefer common population equating to common item equating whenever equivalent random samples are available for linking.

In all NAEP IRT analyses, missing responses at the end of each block a student was administered were considered "not-reached," and treated as if they had not been presented to the respondent. Missing responses before the last observed response in a block were considered intentional omissions, and treated as fractionally correct at the value of the reciprocal of the number of response alternatives. These conventions are discussed by Mislevy and Wu (1988). With regard to the handling of not-reached items, Mislevy and Wu found that ignoring not-reached items introduces slight biases into item parameter estimation to the degree that not-reached items are present and speed is correlated with ability. With regard to omissions, they found that the method described above provides consistent[2] limited-information likelihood estimates of item and ability parameters under the assumption that respondents omit only if they can do no better than responding randomly.

The 3PL pertains to dichotomous, or right/wrong, test items. An increasing number of NAEP tasks, however, yield data in the form of ratings—for example, ratings of 0 through 4 as to quality of the performance. IRT models designed to accommodate such data are currently being examined for implementation in NAEP. Through the 1990 assessment, however, such items comprised relatively small portions of the tasks in the subject areas of reading, mathematics, and science. Ratings for these items were collapsed into dichotomies for IRT analyses.

The average response method (ARM) model. The basic equation of the ARM is an average of item responses:

$$\theta = \underline{a}'\underline{x} . \qquad (11.3)$$

[2] A statistic is a consistent estimate of a parameter if, as sample size increases, the true parameter value is the limit of the expected value of the statistic.

Here $\underline{a}$ is a vector of constants or weights, specified so as to provide a meaningful summary of performance. Weights of 1/n for an n-item test, for example, yield simply an average score; weights given by the k^{th} eigenvector of the covariance matrix for x yield the k^{th} component score. If a respondent responded to all items, then an ARM score would be directly calculable via (11.3) without error. Typically, however, a given NAEP respondent receives only a subset of the items in an ARM scale, so that his or her ARM θ is not observed directly. Extensive information about the ARM as it was used in previous assessments can be found in the 1984, 1986, and 1988 NAEP technical reports.

11.2.2 An Overview of Plausible Values Methodology

Item response theory was developed in the context of measuring individual examinees' abilities. In that setting, each individual is administered enough items (often 60 or more) to permit precise estimation of his or her θ, as a maximum likelihood estimate $\hat{\theta}$, for example. Because the uncertainty associated with each θ is negligible, the distribution of θ, or the joint distribution of θ with other variables, can then be approximated using individuals' $\hat{\theta}$ values as if they were θ values.

This approach breaks down in the assessment setting when, in order to provide broader content coverage in limited testing time, each respondent is administered relatively few items in a scaling area. The problem is that the uncertainty associated with individual θs is too large to ignore, and the features of the $\hat{\theta}$ distribution can be seriously biased as estimates of the θ distribution. (In fact, some students will have response patterns for which the maximum likelihood estimate is infinite—a phenomenon especially problematic in, but not specific to, the assessment setting.) The failure of this approach was verified in early analyses of the 1984 NAEP reading survey; see Wingersky, Kaplan, and Beaton (1987).

In the context of classical test theory, corrections for attenuation transform inconsistent estimates of population characteristics such as variances and correlations to consistent estimates. Simple corrections are not generally available in IRT, but "marginal estimation procedures" (e.g., Mislevy, 1984, 1985) yield consistent estimates of population characteristics directly from item response patterns—without the intermediate step of calculating estimates for each individual. NAEP "plausible values" were developed as a way to provide consistent marginal estimates of key population features, and support approximations of a broad array of additional secondary analyses. More detailed developments of plausible values methodology appear in Mislevy (1991; in press). Along with theoretical justifications, these papers present comparisons with standard procedures, discussions of biases that arise in some secondary analyses, and numerical examples. The following provides a brief overview of the plausible values approach, focusing on its implementation in the 1990 NAEP analyses.

Let $\underline{y}$ represent the responses of all sampled examinees to background and attitude questions, along with design variables such as school membership, and let $\underline{\theta}$ represent the subscale proficiency values. If $\underline{\theta}$ were known for all sampled examinees, it would be possible to compute a statistic $t(\underline{\theta},\underline{y})$—such as a subscale or composite subpopulation sample mean, a sample percentile point, or a sample regression coefficient—to estimate a corresponding population quantity T. A function $U(\underline{\theta},\underline{y})$—e.g., a jackknife estimate—would be used to

gauge sampling uncertainty, as the variance of t around T in repeated samples from the population.

Because the 3PL model is a latent variable model, however, $\underline{\theta}$ values are not observed even for sampled students. To overcome this problem, we follow Rubin (1987) by considering $\underline{\theta}$ as "missing data" and approximate $t(\underline{\theta},\underline{y})$ by its expectation given $(\underline{x},\underline{y})$, the data that actually were observed, as follows:

$$t^*(\underline{x},\underline{y}) = E[t(\underline{\theta},\underline{y})|\underline{x},\underline{y}]$$

$$= \int t(\underline{\theta},\underline{y})\, p(\underline{\theta}|\underline{x},\underline{y})\, d\underline{\theta} \ . \tag{11.4}$$

It is possible to approximate t^* using random draws from the conditional distributions, $p(\underline{\theta}|x_i,y_i)$, of the subscale proficiencies given the item responses x_i and background variables y_i for sampled student i. These values are referred to as "imputations" in the sampling literature, and as "plausible values" in NAEP. The value of $\underline{\theta}$ for any respondent that would enter into the computation of t is thus replaced by a randomly selected value from the conditional distribution $p(\underline{\theta}|x_i,y_i)$. Rubin (1987) proposes that this process be carried out several times—"multiple imputations"—so that the uncertainty associated with imputation can be quantified. The average of the results of, for example, M estimates of t, each computed from a different set of plausible values, is a Monte Carlo approximation of (11.4); the variance among them, B, reflects uncertainty due to not observing θ, and must be added to the estimated expectation of $U(\underline{\theta},\underline{y})$, which reflects uncertainty due to testing only a sample of students from the population. Section 11.3 explains how plausible values are used in subsequent analyses.

It cannot be emphasized too strongly that **plausible values are *not* test scores for individuals** in the usual sense. Plausible values are offered only as intermediary computations for calculating integrals of the form of (11.4), in order to estimate *population* characteristics. When the underlying model is correctly specified, plausible values will provide consistent estimates of population characteristics, even though they are not generally unbiased estimates of the proficiencies of the individuals with whom they are associated. Plausible values differ in a crucial way from the more familiar θ estimates that are in some sense optimal for each examinee (e.g., maximum likelihood estimates, which are consistent estimates of an examinee's θ, and Bayes estimates, which provide minimum mean-squared errors with respect to a reference population): *Point estimates that are optimal for individual examinees have distributions that can produce decidedly nonoptimal (specifically, inconsistent) estimates of population characteristics* (Little & Rubin, 1983). Plausible values, on the other hand, are constructed explicitly to provide consistent estimates of population effects.

11.2.3 Computing Plausible Values in IRT-based Scales

Plausible values for each respondent i are drawn from the conditional distribution $p(\theta|x_i,y_i)$. This subsection describes how, in IRT-based scales, these conditional distributions

are characterized, and how the draws are taken. An application of Bayes' theorem with the IRT assumption of conditional independence produces

$$p(\underline{\theta}|x_i,y_i) \propto P(x_i|\underline{\theta},y_i)\, p(\underline{\theta}|y_i)$$

$$= P(x_i|\underline{\theta})\, p(\underline{\theta}|y_i), \tag{11.5}$$

where, for vector-valued $\underline{\theta}$, $P(x_i|\underline{\theta})$ is the product over subscales of the *independent likelihoods* induced by responses to items within each subscale, and $p(\underline{\theta}|y_i)$ is the multivariate—and generally nonindependent—*joint density* of proficiencies for the subscales, conditional on the observed value y_i of background responses.

In the analyses of the data from the 1990 NAEP assessment, a normal (Gaussian) form was assumed for $p(\underline{\theta}|y_i)$, with a common dispersion and with a mean given by a linear model based on selected main-effects and two-way interactions of the complete vector of background variables. The included background variables will be referred to as the *conditioning variables*, and will be denoted y^c. The effects (conditioning variable contrasts) are listed in Appendix F. In the mathematics and science cross-sectional analyses, the leading principal components of the effects (sufficient to account for 90 percent of their total variance) were employed, in order to enhance the stability of the solution in models with a larger number of effects; see Chapters 13 and 14 for details. The following model was fit to the data within each age/grade cohort for a given subject area:

$$\theta = \Gamma' y^c + \varepsilon \, , \tag{11.6}$$

where ε is normally distributed with mean zero and dispersion Σ. As in regression analysis, Γ is a vector or matrix in which each column contains the *effects* for one subscale and Σ is the scalar or matrix *variance of residuals*. By fitting the model (11.6) separately within each age/grade cohort, interactions between each cohort and the conditioning variables are automatically included in the conditional joint density of subscale proficiencies. Like item parameter estimates, the estimates of the parameters of conditional distributions were treated as known true values in subsequent steps of the analyses.

Maximum likelihood estimates of Γ and Σ were obtained with Rogers' (1991) enhanced version of Sheehan's (1985) MGROUP computer program, using a variant of the EM solution described in Mislevy (1985). The difference from the published algorithm lies in the numerical approximation that was employed, which we now describe. Note from (11.5) that $p(\underline{\theta}|x_i,y_i)$ is proportional to the product of two terms, the likelihood $P(x_i|\underline{\theta})$ and the conditional distribution $p(\underline{\theta}|y_i)$. The conditional distribution for person i has been assumed normal (multivariate normal, when there are multiple subscales), with mean $\mu_i^c = \Gamma' y_i^c$ and

covariance matrix Σ; if the likelihood is approximated by another normal distribution, with mean μ_i^L and covariance matrix Σ_i^L; then the posterior $p(\underline{\theta}|x_i,y_i)$ is also normal with covariance matrix

$$\Sigma_i^P = (\Sigma^{-1} + (\Sigma_i^L)^{-1})^{-1} \tag{11.7}$$

and mean

$$\bar{\underline{\theta}}_i = (\underline{\theta}_i^c \Sigma^{-1} + \underline{\theta}_i^L (\Sigma_i^L)^{-1})(\Sigma_i^P). \tag{11.8}$$

In the analyses of the 1990 NAEP assessment, a normal approximation for $P(x_i|\underline{\theta})$ is accomplished in a given scale by the steps described below. (Recall that by the assumed conditional independence across scales, the joint conditional likelihood for multiple scales is the product of independent likelihoods for each of the scales.) These computations are carried out in the scale determined by NAEP-BILOG item parameter estimates, where the provisional mean and standard deviation of the composite population formed by combining the three NAEP age/grades has mean zero and standard deviation one. The steps were as follows.

1) Lay out a grid of Q equally spaced points from -5 to +5, a range that covers the region in each scale where all examinees are virtually certain to occur. The value of Q varies from 20 to 40, depending on the subscale being used; smaller values suffice for subscales with few items given to each respondent, while larger values are required for subscales with many items.

2) At each point X_q, compute the likelihood $L(x_i|\theta=X_q)$.

3) To improve the normal approximation in those cases in which likelihoods are not approximately symmetric in the range of interest—as when all of a respondent's answers are correct—multiply the values from Step 2 by the mild smoothing function

$$S(X_q) = \frac{\exp(X_q+5)}{[1+\exp(X_q+5)][1+\exp(X_q-5)]} \tag{11.9}$$

This is equivalent to augmenting each examinee's response vector with responses to two fictitious items, one extraordinarily easy item that everyone gets right and one extraordinarily difficult item that everyone gets wrong. This expedient improves the normal approximation for examinees with flat or degenerate likelihoods in the range where their conditional distributions lie, but has negligible effects for examinees with even modestly well-determined symmetric likelihoods.

4) Compute the mean and standard deviation of θ using the weights $S(X_q)L(x_i|\theta=X_q)$ obtained in Step 3.

At this stage the likelihood induced by a respondent's answers to the items in a given scale is approximated by a normal distribution. Since the mathematics and science cross-sectional analyses use multiple subscales, independent normal distributions, one per subscale, are used to summarize information from responses to items from the several subscales.

This normalized-likelihood/normal posterior approximation was then employed in both the estimation of Γ and Σ and in the generation of plausible values. From the final estimates of Γ and Σ, a respondent's posterior distribution was obtained from the normal approximation using the four-step procedure outlined above. A plausible value—vector-valued for multiple subscale analyses—was drawn at random from this normal distribution. Finally, in multiple subscale analyses, weighted-average composites over subscales were also calculated after appropriate rescaling (see the appropriate subject area chapters for specifications of rescaling conventions and definitions of composites).

11.3 ANALYSES

When survey variables are observed without error from every respondent, standard formulae are available for calculating statistics that estimate population characteristics, and for associated variance estimators (e.g., NAEP's jackknife) that quantify the uncertainty associated with sample statistics due to the sampling of respondents from the population. Item percents correct for NAEP cognitive items meet this requirement, but scale-score proficiency values do not. The fact that θ values are not observed even for the respondents in the sample requires additional statistical analyses to draw inferences about θ distributions and to quantify the uncertainty associated with those inferences. As described above, Rubin's (1987) multiple imputations procedures were adapted to the context of latent variable models to produce the plausible values upon which many analyses of the data from the 1990 NAEP assessment were based. This section describes how plausible values were employed in subsequent analyses to yield inferences about population and subpopulation distributions of proficiencies.

11.3.1 Computational Procedures

Suppose one wishes to draw inferences about a number $T(\underline{\theta},\underline{Y})$ that could be calculated explicitly if the θ and y values of each member of the population were known. Suppose further that if θ values were observable, we would be able to estimate T from a sample of N pairs of θ and y values by the statistic $t(\underline{\theta},\underline{y})$ [where $(\underline{\theta},\underline{y}) = (\theta_1,y_1,\ldots,\theta_N,y_N)$], and that we could estimate the variance in t around T due to sampling respondents by the function $U(\underline{\theta},\underline{y})$. Given that observations consist of (x_i,y_i) rather than (θ_i,y_i), we can approximate t by its expected value conditional on $(\underline{x},\underline{y})$, or

$$\begin{aligned} t^*(\underline{x},\underline{y}) &= E[t(\underline{\theta},\underline{y})\,|\,\underline{x},\underline{y}] \\ &= \int t(\underline{\theta},\underline{y})\, p(\underline{\theta}\,|\,\underline{x},\underline{y})\, d\underline{\theta}\,. \end{aligned} \qquad (11.10)$$

It is possible to approximate t^* with random draws from the conditional distributions $p(\underline{\theta}_i | x_i, y_i)$, which are obtained for all respondents by the method described in section 11.3.3. Let $\hat{\underline{\theta}}_m$ be the m^{th} such vector of "plausible values," consisting of a (possibly multidimensional) value for the latent variable of each respondent. This vector is a plausible representation of what the true $\underline{\theta}$ vector might have been, had we been able to observe it.

The following steps describe how an estimate of a scalar statistic $t(\underline{\theta}, \underline{y})$ and its sampling variance can be obtained from M (> 1) such sets of plausible values. (Five sets of plausible values are used in NAEP analyses.)

1) Using each set of plausible values $\hat{\underline{\theta}}_m$ in turn, evaluate t as if the plausible values were true values of $\underline{\theta}$. Denote the results $\hat{t}_m$, for $m = 1, \ldots, M$.

2) Using the jackknife variance estimator defined in Chapter 10, compute the estimated sampling variance of $\hat{t}_m$, denoting the result U_m, for $m = 1, \ldots, M$.

3) The final estimate of t is

$$t^* = \sum_{m=1}^{M} \frac{\hat{t}_m}{M}. \tag{11.11}$$

4) Compute the average sampling variance over the M sets of plausible values, to approximate uncertainty due to sampling respondents:

$$U^* = \sum_{m=1}^{M} \frac{U_m}{M}. \tag{11.12}$$

5) Compute the variance among the M estimates $\hat{t}_m$, to approximate uncertainty due to not observing θ values from respondents:

$$B_M = \sum_{m=1}^{M} \frac{(\hat{t}_m - t^*)^2}{(M - 1)} \tag{11.13}$$

6) The final estimate of the variance of t^* is the sum of two components:

$$V = U^* + (1 + M^{-1})\, B_M. \tag{11.14}$$

Note: Due to the excessive computation that would be required, NAEP analyses did not compute and average jackknife variances over all five sets of plausible values, but only on the first set. Thus, in NAEP reports, U^* is approximated by U_1.

11.3.2 Statistical Tests

Suppose that if θ values were observed for sampled students, the statistic $(t - T)/U^{1/2}$ would follow a t-distribution with d degrees of freedom. Then the incomplete-data statistic $(t^* - T)/V^{1/2}$ is approximately t-distributed, with degrees of freedom given by

$$\nu = \frac{1}{\dfrac{f_M^2}{M-1} + \dfrac{(1-f_M)^2}{d}}$$

where f_M is the proportion of total variance due to not observing θ values:

$$f_M = (1+M^{-1})\, B_M / V_M \ . \qquad (11.15)$$

When B_M is small relative to U^*, the reference distribution for incomplete-data statistics differs little from the reference distribution for the corresponding complete-data statistics. This is the case with main NAEP reporting variables. If, in addition, d is large, the normal approximation can be used to flag "significant" results.

For k-dimensional t, such as the k coefficients in a multiple regression analysis, each U_m and U^* is a covariance matrix, and B_M is an average of squares and cross-products rather than simply an average of squares. In this case, the quantity

$$(T-t^*)\, V^{-1}\, (T-t^*)' \qquad (11.16)$$

is approximately F distributed, with degrees of freedom equal to k and ν, with ν defined as above but with a matrix generalization of f_M:

$$f_M = (1+M^{-1})\ \mathrm{Trace}\ (B_M V_M^{-1})/k. \qquad (11.17)$$

By the same reasoning as used for the normal approximation for scalar t, a chi-square distribution on k degrees of freedom often suffices.

11.3.3 Biases in Secondary Analyses

Statistics t^* that involve proficiencies in a scaled content area and variables included in the conditioning variables y^c, are consistent estimates of the corresponding population values T. Statistics involving background variables y that were *not* conditioned on, or relationships among proficiencies from *different* content areas, are subject to asymptotic biases whose magnitudes depend on the type of statistic and the strength of the relationships of the nonconditioned background variables to the variables that were conditioned on and to the proficiency of interest. That is, the large sample expectations of certain sample statistics need not equal the true population parameters.

The *direction* of the bias is typically to underestimate the effect of nonconditioned variables. For details and derivations see Beaton and Johnson (1990), Mislevy (1991), and

Mislevy and Sheehan (1987, section 10.3.5). For a given statistic t^* involving one content area and one or more nonconditioned background variables, the *magnitude* of the bias is related to the extent to which observed responses x account for the latent variable θ, and the degree to which the nonconditioned background variables are explained by conditioning background variables. The first factor—conceptually related to test reliability—acts consistently in that greater measurement precision reduces biases in *all* secondary analyses. The second factor acts to reduce biases in certain analyses but increase it in others. In particular,

- High shared variance between conditioned and nonconditioned background variables *mitigates* biases in analyses that involve only proficiency and nonconditioned variables, such as marginal means or regressions.

- High shared variance *exacerbates* biases in regression coefficients of conditional effects for nonconditioned variables, when nonconditioned and conditioned background variables are analyzed jointly as in multiple regression.

The large number of background variables that have been included in the conditioning vector for the 1990 NAEP assessments allows a large number of secondary analyses to be carried out with little or no bias, and mitigates biases in analyses of the marginal distributions of θ in nonconditioned variables. Kaplan and Nelson's analysis of the 1988 NAEP reading data (some results of which are summarized in Mislevy, 1991), which had a similar design and fewer conditioning variables, suggest that the potential bias for nonconditioned variables in multiple regression analyses is in the neighborhood of 10 percent, and biases in simple regression of such variables, 5 percent. Additional research (summarized in Mislevy, 1991) indicates that most of the bias reduction obtainable from conditioning on a large number of variables can be captured by instead conditioning on the first several principal components of the matrix of all original conditioning variables. This procedure was adopted for the 1990 mathematics and science cross-sectional assessments, by replacing the original conditioning effects by the first K principal components, where K was selected so that 90 percent of the total variance of the full set of conditioning variables (after standardization) was captured. Mislevy (1991) shows that this puts an upper bound of 10 percent on the average potential bias for analyses involving the original conditioning variables.

11.3.4 A Numerical Example

To illustrate how plausible values are used in subsequent analyses, this subsection gives some of the steps in the calculation of 1988 grade-level reading means and their estimation-error variances.

The weighted mean of the first plausible values of the grade 4 students in the sample is 230.68, and the jackknife variance of these values is 1.17. Were these values true θ values, then 230.68 would be the estimate of the mean and 1.17 would be the estimation-error variance. The weighted mean of the second plausible values of the same students, however, is 230.60; the third, fourth, and fifth plausible values give weighted means of 230.19, 230.32, and 230.06. Since all of these figures are based on precisely the same sample of students, the variation among them is due to uncertainty about the students' θs, having observed their item responses and background variables. Taking the jackknife variance estimate from the first

plausible value, 1.17, as our estimate U^* of sampling variance, and the variance among the five weighted means, .09, as our estimate B of uncertainty due to not observing θ, we obtain as the final estimate V of total error variance $1.17 + (1+5^{-1})\ .09 = 1.28$.

With U^* and B defined as above, and with $M=5$, we may obtain a value for Rubin's (1987) index characterizing the relative increase in variance due to the latency of θ:

$$r = (1+M^{-1})B_M/U^* .$$

In grade 4, $r = .09$. Corresponding values were also calculated for grade 8 and grade 12. The results are shown in Table 11-1. It is also possible to partition the estimation error variance of a statistic using these same variance components. The proportion of error variance due to sampling students from the population is U^*/V, and the proportion due to the latent nature of θ is $(1+M^{-1})B_M/V$.

Table 11-1
Estimation Error Variance and Related Coefficients
for the 1988 Grade-level Reading Assessments

					Proportion of Variance Due to...	
Grade	U^*	$(1+5^{-1})B$	V	r	Student Sampling: U^*/V	Latency of θ: $(1+5^{-1})B/V$
4	1.17	.11	1.28	.09	0.91	0.09
8	.96	.07	1.03	.07	0.93	0.07
12	.69	.02	.71	.03	0.97	0.03

11.4 OVERVIEW OF THE 1990 NAEP SCALES

IRT scale-score analyses were carried out in the following subject areas in the 1990 NAEP assessment.

- Reading: One IRT scale linking 1990 results to the trend line established in the 1984 reading assessment, and one IRT scale linking 1990 results to the 1988 assessment.

- Mathematics: Five newly developed subscales for cross-sectional analyses, and a unidimensional IRT mathematics scale linked to the 1986 mathematics assessment for trend analyses.

- Science: Four newly developed subscales for cross-sectional analyses, and a unidimensional science scale linked to the 1986 science assessment. The arbitrary origins and unit-sizes of the cross-sectional subscales were matched to composite results of the 1986 science assessment, which was built around different age definitions, objectives, and item pools.

Details follow in Chapters 12, 13, and 14.

Chapter 12

DATA ANALYSIS FOR THE READING ASSESSMENT[1]

John R. Donoghue

Educational Testing Service

This chapter describes the analyses performed on the responses to the cognitive and background items in the 1990 assessment of reading. These analyses led to the results presented in *Trends in Academic Progress: Achievement of U.S. Students in Science, 1969-70 to 1990; Mathematics, 1973 to 1990; Reading, 1971 to 1990; and Writing, 1984 to 1990* (Mullis, Dossey, Foertsch, Jones, & Gentile, 1991) and in *Reading in School and out of School: Students' Literary Experience and Academic Achievement from 1988 to 1990 at Grades 4, 8, and 12* (Foertsch, 1992). The emphasis of this chapter is on the methods and results of procedures used to develop the IRT-based scale scores that formed the basis of these reports. However, some attention is given to the analysis of open-ended items as reported in *Reading in School and out of School.* The theoretic underpinnings of the IRT and plausible value methodology described in this chapter are given in Chapter 11.

The objectives of the reading analyses were to

- prepare scale values and perform all analyses necessary to produce a long-term trend report in reading. The reading trend results include the years 1971, 1975, 1980, 1984, 1988, and 1990.
- prepare scale values for the cross-sectional analysis of the main focused-BIB reading samples.
- if feasible, link the 1990 main focused-BIB samples to the 1988 scale. Perform all analyses necessary to produce the short-term trend. The short-term trend includes only the years 1988 and 1990.

The student samples that were administered reading items in the 1990 assessment are shown in Table 12-1. (See Chapters 1 and 3 for descriptions of the target populations and the sample design used for the assessment.) Data from the first four samples (denoted Rdg-MainP and Rdg-ABB) were used in the cross-sectional analysis and formation of the short-term trend, while the rest of the samples were collected for long-term trend analyses.

[1] Data analysis and scaling were performed by Drew Bowker, Minhwei Wang, and David Freund. Rebecca Zwick oversaw analysis of the trend data. Nancy Allen, Robert Mislevy, and Kentaro Yamamoto consulted on IRT scaling and generation of plausible values.

The cross-sectional data from the 1990 main focused-BIB samples were scaled separately from the data from the bridge samples that contributed to the trends in reading achievement. Accordingly, the trend and cross-sectional analyses are presented in separate sections. Section 12.1 pertains to the scaling of the data from the trend bridges; section 12.2 contains information about the scaling of the data from the main focused-BIB samples and the formation of the short-term trend.

Table 12-1
NAEP 1990 Reading Student Samples

Sample	Booklets	Mode	Cohort Assessed	Time of Testing	Age Definition	Modal Grade	Number Assessed
9 [Rdg-MainP]	1-7	Print	Age 9/grade 4	Winter, spring	CY	4	8,480
13 [Rdg-MainP]	1-7	Print	Age 13/grade 8	Winter, spring	CY	8	8,725
17 [Rdg-MainP]	1-7	Print	Age 17/grade 12	Winter, spring	CY	12	8,351
9 [Rdg-ABB]	8-10	Print	Age 9/grade 4	Winter, spring	CY	4	3,615
9 [RW-Br84]	51-56	Print	Age 9/grade 4	Winter	CY	4	5,926
13 [RW-Br84]	51-56	Print	Age 13/grade 8	Fall	CY	8	6,233
17 [RW-Br84]	51-56	Print	Age 17/grade 11	Spring	Not CY	11	5,614
9 [RMS-Br86]	91-93	Mixed	Age 9	Winter	CY	4	6,235
13 [RMS-Br86]	91-93	Mixed	Age 13	Fall	CY	8	6,649
17 [RMS-Br86]	61-66	Print	Age 17/grade 11	Spring	Not CY	11	8,338

LEGEND:

Rdg	Reading
RW	Reading and writing
RMS	Reading, mathematics, and science
MainP	Main assessment, print administration
ABB	Answer book bridge (main assessment)
Br84	Bridge to 1984
Br86	Bridge to 1986
Print	Printed administration
Mixed	Mathematics and science administered by audiotape, reading administered by print
CY	Calendar year: birthdates in 1980, 1976, and 1972 for ages 9, 13, and 17
Not CY	Age 17 only: birthdates between Oct. 1, 1972 and Sept. 30, 1973

12.1 TREND DATA ANALYSIS

The trend results reported in *Trends in Academic Progress* are based on print administrations and occur at all of the age levels. The samples involved in the analysis are shown in Table 12-1 as 9[RW-Br84], 13[RW-Br84], and 17[RW-Br84]. The bridge booklets for these samples contained blocks of reading and writing items. All items were presented in print form. All students received a block of common background questions, distinct for each age, in addition to subject-area background questions, which were presented in the cognitive blocks. The booklets are identical to those used for bridge assessments in 1984 and 1988. The booklets and the blocks within those booklets are listed in Tables 4-7, 4-8, and 4-9 of Chapter 4. Additional information about all of the items in these blocks is in Tables 4-10, 4-11, and 4-12 of

that chapter. This chapter includes specific information about the trend items that were scaled. Both age- and grade-selected students contributed to the trend scaling. However, only students in the "age-only" portion of the reading trend samples contributed to the results presented in *Trends in Academic Progress.*

Table 12-2 clarifies the relationships between the 1990 trend samples and samples from previous years. For ages 9, 13 and 17, the [RW-Br84] samples allow direct comparisons with 1988 samples, as well as with 1984 samples. The current trend scale was established in 1984. The 1971, 1975, and 1980 assessments were linked to the 1984 assessment using a complex equating strategy, which is described in *Implementing the New Design: The NAEP 1983-84 Technical Report* (Beaton, 1987). At each age, several intact booklets were retained from the 1984 assessment. These intact booklets form the basis of the reading trend assessment in 1988 and 1990. Information about the 1988 assessment is available in *Focusing the New Design: The NAEP 1988 Technical Report* (Johnson & Zwick, 1990).

In addition to the samples that contributed data to *Trends in Academic Progress*, another type of trend sample [RMS-Br86] was collected in 1990. These data were collected for two reasons. First, the reading blocks in the sample appeared in booklets also containing blocks that form the basis of the long-term trends in mathematics and in science. By maintaining intact booklets, NAEP preserves the context for the mathematics and science trend assessments. The second reason for gathering the [RMS-86] sample involved the investigation of the 1986 "reading anomaly." Preliminary results for the 1986 reading assessment revealed large decreases from the 1984 performance of 9- and 17-year-olds, while the performance of 13-year-olds was found to have increased slightly. It was deemed unlikely that such large changes could have taken place in such a short time. Thus, reporting of the 1986 results was delayed until subsequent research could determine whether the results represented true changes in reading ability. Consequently, a special study was included in the 1988 assessment, the results of which are summarized in *The Effect of Changes in the National Assessment: Disentangling the NAEP 1985-86 Reading Anomaly* (Beaton & Zwick, 1990). The bridge to 1986 is a sample that is comparable to an assessment of students that was part of the 1988 study of the 1986 reading anomaly. The booklets, presentation, time of assessment, and age definition were the same as those for the main NAEP sample in 1986. At the time the design of the 1990 was finalized, the study of the reading anomaly had not been completed, and one purpose in including the [RMS-Br86] samples was to gather additional data for investigation. The original data from the reading anomaly study proved to be sufficient, but only after the 1990 data collection had begun. The results from the [RMS-Br86] sample were not included in *Trends in Academic Progress*, and detailed analyses of these data were not conducted.

The 1990 bridge to 1984 included, at each age level, six of the assessment booklets administered in 1984. These booklets (51-56) contained both reading and writing blocks, as well as background items. Although these bridge booklets represented only about a tenth of the reading booklets administered in the complex 1984 BIB design,[2] they contained 10 of the 12 reading blocks that were scaled at each age/grade level in 1984. The samples of students who

[2] The bridge to 1984 included 1984 booklets 16, 17, 27, 34, 55, and 60 at age 9 and booklets 13, 16, 17, 21, 34, and 57 at ages 13 and 17 (see J. R. Johnson, 1987, pp. 120-121). The 1984 BIB assessment included 57 booklets that contained at least one scaled reading block at age 9 and 56 such booklets at ages 13 and 17.

Table 12-2
NAEP Reading Samples Contributing to 1990 Trend Results, 1971-1990

Cohort	Year	Sample	Subjects	Time of Testing	Mode of Administration	Age Definition	Modal Grade
Age 9	1971	Main	RL	Winter	Tape	CY	4
	1975	Main	RA	Winter	Tape	CY	4
	1980	Main	RA	Winter	Tape	CY	4
	1984	Main	RW	Winter, spring	Print	CY	4
	1984	BrLT	RW	Winter	Tape	CY	4
	1988	Br84	RW	Winter	Print	CY	4
	1990	Br84	RW	Winter	Print	CY	4
Age 13	1971	Main	RL	Fall	Tape	CY	8
	1975	Main	RA	Fall	Tape	CY	8
	1980	Main	RA	Fall	Tape	CY	8
	1984	Main	RW	Winter, spring	Print	CY	8
	1984	BrLT	RW	Fall	Tape	CY	8
	1988	Br84	RW	Fall	Print	CY	8
	1990	Br84	RW	Fall	Print	CY	8
Age 17	1971	Main	RL	Spring	Tape	Not CY	11
	1975	Main	RABS	Spring	Tape	Not CY	11
	1980	Main	RA	Spring	Tape	Not CY	11
	1984	Main	RW	Winter, spring	Print	Not CY	11
	1984	BrLT	RW	Spring	Tape	Not CY	11
	1988	Br84	RW	Spring	Print	Not CY	11
	1990	Br84	RW	Spring	Print	Not CY	11

LEGEND:

RL Reading and literature
RA Reading and art
RABS Reading, art, index of basic skills
RW Reading and writing

Main Main assessment
BrLT Bridge for long-term trend
Br84 Bridge to 1984 (these samples received common booklets within an age group)

Print Print administration
Tape Audiotape administration

CY Calendar year: birthdates (1990 sample) in 1980 and 1976 for ages 9 and 13

Not CY Age 17 only (1990 sample): birthdates between Oct 1, 1972 and Sept. 30, 1973

received these bridge booklets are described in Table 12-1 and in Chapter 4. The purpose of the long-term reading trend analysis was to add to the reading trend results that extended from 1971 to 1988 for ages 9, 13, and 17. The numbers of scaled items for each age are presented in Table 12-3. Each age was scaled separately. The numbers of items scaled in 1990 that were common across assessment years are given in Table 12-4. As was the case for previous trend analyses, the trend scale is univariate. Dimensionality analyses conducted following the 1984 assessment showed that the reading items were well summarized by a unidimensional scale (Zwick, 1987a).

The steps in the reading trend analysis are documented in the following sections. As is usual in NAEP analyses, the first step was to gather item and block information. Next, the trend items were calibrated and plausible values were generated after conditioning on available background variables. Finally, the scale values were placed on the final reading trend proficiency scale used in previous trend assessments.

12.1.1 Item Analysis for the Bridge-to-1984 Assessment

Conventional item analyses did not identify any difficulties with the bridge data for the samples that bridge to 1984. Table 12-5 contains the number of items, size of the sample administered the block, mean weighted proportion correct, mean weighted r-biserial, and mean weighted alpha as a measure of reliability for each block. Because the blocks were presented in self-paced, print form, the weighted proportion of students attempting the last item is included in the table to give an indication of the speededness of each block. Common labeling of these blocks across ages does not denote common items. Student weights were used for all statistics, except for the sample sizes. The average values reflect only the items in the block that were scaled. The 1990 item-level statistics were not very different from those for the 1984 and 1988 assessments.

12.1.2 Treatment of Open-ended Items

Data for open-ended items in the trend analysis were used for the 1984 and 1990 assessments only. Open-ended items were not included in the original scoring of the 1988 reading assessment because a previous study (Zwick, 1988) had shown that scoring inconsistencies had affected these items. Therefore, 1988 data for the open-ended items were not used, in order to be consistent with the original scaling of the data. A similar review was performed on the 1990 open-ended items. In general, the 1990 scoring did not suffer from the same inconsistencies as the 1988 scoring. Therefore, the 1990 open-ended items were used in the trend analysis.

At each age, several open-ended items were found to exhibit drops in interrater reliability and/or scorer drift—that is, the professional scorers who rated the current assessment showed evidence of rating items more strictly or more leniently than did scorers for the 1984 assessment. Items exhibiting marked item drift were excluded from calibration. Eight of the 198 total trend reading items were excluded. These items are listed in Table 12-6. The remaining open-ended items were dichotomized according to criteria developed by subject-area

Table 12-3

Numbers of Scaled Reading Trend Items Common Across Ages

Age	Number of Items
9 only	62
13 only	21
17 only	24
9 and 13 only	13
9 and 17 only	2
13 and 17 only	44
9, 13, and 17	24
Total	190

Table 12-4

Numbers of Scaled Reading Trend Items Common Across Assessments

Assessment Year	Number of Items		
	Age 9	Age 13	Age 17
1984, 1990	101	102	94
1984, 1988, 1990	98	99	87
1980, 1984, 1988, 1990	67	72	52
1975, 1980, 1984, 1988, 1990	36	46	37
1971, 1975, 1980, 1984, 1988, 1990	36	46	37

Table 12-5

Descriptive Statistics for Item Blocks
Reading Trend Samples

Statistics	Blocks										
	BH	BJ	BK	BL	BM	BN	BO	BP	BQ	BR	BV
Age 9											
Number of scaled items	9	8	11	7	11	12	11	—	11	12	9
Number of scaled open-ended items	0	0	0	1	1	1	0	—	0	0	3
Unweighted sample size	723	720	721	686	694	682	694	—	689	1422	711
Average weighted proportion correct	.65	.51	.41	.50	.40	.56	.47	—	.54	.47	.62
Average weighted r-biserial	.78	.73	.67	.82	.66	.76	.59	—	.72	.67	.75
Weighted alpha reliability	.76	.70	.74	.75	.73	.83	.62	—	.79	.77	.76
Weighted proportion of students attempting last item	.99	.94	.84	.76	.66	.72	.90	—	.88	.89	.96
Age 13											
Number of scaled items	11	9	8	5	10	12	10	9	17	11	—
Number of scaled open-ended items	0	0	0	0	0	1	1	1	0	0	—
Unweighted sample size	738	766	758	787	758	774	783	761	787	736	—
Average weighted proportion correct	.69	.62	.63	.72	.61	.64	.64	.66	.58	.68	—
Average weighted r-biserial	.67	.65	.74	.88	.69	.66	.60	.69	.54	.72	—
Weighted alpha reliability	.61	.61	.68	.60	.67	.75	.50	.61	.70	.74	—
Weighted proportion of students attempting last item	.98	.93	.99	.98	.96	.77	.86	.89	.80	.98	—
Age 17											
Number of scaled items	12	5	8	5	11	12	13	10	11	7	—
Number of scaled open-ended items	1	1	0	0	1	1	1	1	1	0	—
Unweighted sample size	737	739	723	722	725	728	727	734	716	729	—
Average weighted proportion correct	.71	.60	.79	.84	.68	.81	.66	.72	.57	.69	—
Average weighted r-biserial	.77	.74	.85	.93	.78	.77	.61	.67	.59	.83	—
Weighted alpha reliability	.72	.34	.67	.43	.69	.73	.72	.66	.63	.73	—
Weighted proportion of students attempting last item	.97	.97	1.00	1.00	.96	.89	.62	.85	.73	.99	—

experts. The dichotomized versions of the open-ended items were included in the calibration. However, to further ensure stability of the common calibration, these items were scaled separately for each administration, because the open-ended items were not scored for the 1988 assessment.

Table 12-6
Items Deleted from the Reading Trend Analysis

Age	Block	Item	Reason for Exclusion
9	BH	N001507	Exclude, marked score drift
	BJ	N001801	Never scaled, extremely low probability of correct response
	BM	N003003	Exclude, poor fit to IRT model
	BJ	N008905	Exclude, marked score drift
13	BH	N001507	Exclude, marked score drift
	BJ	N001904	Exclude, marked score drift
	BK	N002302	Never scaled, nonordinal item
	BL	N002804	Exclude, marked score drift
	BM	N003104	Exclude, drop in interrater agreement
17	BK	N002302	Never scaled, nonordinal item
	BL	N002804	Exclude, score drift

12.1.3 Estimation of Item Parameters

The first step in the scaling process was the estimation of item parameters for the trend items. This item calibration was performed using the NAEP-BILOG program (Rogers & Nelson, 1990; Mislevy & Bock, 1982). Items were calibrated separately for each of the three age groups. Item parameters were estimated using combined data from the assessment years 1984, 1988 and 1990, treating each assessment as a sample from a separate subpopulation. The calibration was performed on a subsample of all the available subjects, with 1984 and 1990 assessments sampled down to have the same (unweighted) number of subjects as the 1988 assessment. This resulted in approximately 300-500 examinees in each assessment year for each item. Student weights were used for the analysis.

At each age, several items were calibrated separately for each assessment year, for one of two reasons. First, as noted above, all open-ended items that were included in scaling were calibrated separately across assessments. Second, empirical item response functions (IRFs) for each assessment were compared to the common IRF fit by NAEP-BILOG. Overall, 42 of then 190 trend reading items were calibrated separately by assessment year. Table 12-7 shows the items for which empirical IRFs differed across assessments. These items were temporarily removed from the calibration, and a preliminary scaling was performed using only items which exhibited good fit to the common IRF. A second calibration was then performed separately for each assessment. This calibration held constant the parameters of items used in the first calibration while estimating parameters for those items to be calibrated separately across assessments (the items listed in Table 12-7). One item (N003003) was deleted from the scaling of the age 9 trend data, due to poor fit to the IRT model; the item was too difficult for these

Table 12-7
Items Calibrated Separately by Assessment Year in the Reading Trend Analysis

Age	Block	Item	Reason for Separate Calibration
9	BH	N001101	Poor fit across assessments to common IRF
	BJ	N001603	Poor fit across assessments to common IRF
	BK	N002002	Poor fit across assessments to common IRF
	BL	N002804	Open-ended
	BM	N003002	Poor fit across assessments to common IRF
	BM	N003101	Poor fit across assessments to common IRF
	BM	N003104	Open-ended
	BN	N003704	Open-ended
	BH	N009002	Poor fit across assessments to common IRF
	BK	N009003	Poor fit across assessments to common IRF
	BQ	N010903	Poor fit across assessments to common IRF
	BN	N014302	Poor fit across assessments to common IRF
	BV	N014502	Poor fit across assessments to common IRF
13	BH	N001101	Poor fit across assessments to common IRF
	BH	N001201	Poor fit across assessments to common IRF
	BH	N001301	Poor fit across assessments to common IRF
	BJ	N001701	Poor fit across assessments to common IRF
	BK	N002101	Poor fit across assessments to common IRF
	BL	N002701	Poor fit across assessments to common IRF
	BN	N003704	Open-ended
	BO	N003801	Poor fit across assessments to common IRF
	BO	N004301	Poor fit across assessments to common IRF
	BO	N004303	Open-ended
	BP	N004501	Poor fit across assessments to common IRF
	BP	N004601	Poor fit across assessments to common IRF
	BP	N004605	Open-ended
	BQ	N005203	Poor fit across assessments to common IRF
	BQ	N005303	Poor fit across assessments to common IRF
	BR	N005403	Poor fit across assessments to common IRF
	BR	N005406	Poor fit across assessments to common IRF
17	BH	N001502	Poor fit across assessments to common IRF
	BH	N001504	Poor fit across assessments to common IRF
	BH	N001507	Open-ended
	BJ	N001702	Poor fit across assessments to common IRF
	BJ	N001703	Poor fit across assessments to common IRF
	BJ	N001904	Open-ended
	BK	N002003	Poor fit across assessments to common IRF
	BM	N002904	Poor fit across assessments to common IRF
	BM	N003001	Poor fit across assessments to common IRF
	BM	N003104	Open-ended
	BN	N003501	Poor fit across assessments to common IRF
	BN	N003704	Open-ended
	BO	N003801	Poor fit across assessments to common IRF
	BO	N004303	Open-ended
	BP	N004501	Poor fit across assessments to common IRF
	BP	N004602	Poor fit across assessments to common IRF
	BP	N004605	Open-ended
	BQ	N005002	Poor fit across assessments to common IRF
	BQ	N005003	Poor fit across assessments to common IRF
	BQ	N005202	Poor fit across assessments to common IRF
	BN	N015201	Poor fit across assessments to common IRF
	BQ	N015903	Poor fit across assessments to common IRF
	BQ	N015905	Open-ended
	BO	N016004	Poor fit across assessments to common IRF
	BH	N017003	Poor fit across assessments to common IRF

students to yield reliable estimates of item parameters. No other trend items were deleted from the 1990 trend analysis. A list of the items scaled for each of the ages, along with their item parameter estimates, appears in Tables E-8, E-9, and E-10 in Appendix E.

12.1.4 Generation of Plausible Values

The generation of plausible values was conducted independently by age for each of the three assessment years. The item parameters from NAEP-BILOG, final student weights, item responses, and selected background variables were used with the NAEP version (Rogers, 1991) of the computer program MGROUP (Sheehan, 1985) to generate the values for each age. There were 35 contrasts in the conditioning model at age 9, 35 at age 13, and 33 at age 17. Appendix F gives the codings for the conditioning variables (Table F-3) and the estimated conditioning effects (Tables F-12, F-13, and F-14) for the three age groups. The estimated conditioning effects in the tables are expressed on the scale of the original calibration.

12.1.5 The Final Proficiency Scale

The linear indeterminacy of the trend scale was resolved by linking the 1990 trend scales to previous trend scales. For each age, the item parameters from 1990 based on data from 1984, 1988, and 1990 were used with the 1984 data to find plausible values for the 1984 data. The mean and standard deviation of all of the plausible values was calculated and matched to the mean and standard deviation of all of the plausible values based on the 1984 item parameters and 1984 data as given in earlier reports. The transformations that resulted from this matching of the first two moments for the 1984 data are

Age 9: $\theta_{proficiency} = 42.28 \bullet \theta_{calibrated} + 239.84$

Age 13: $\theta_{proficiency} = 36.24 \bullet \theta_{calibrated} + 246.81$

Age 17: $\theta_{proficiency} = 42.12 \bullet \theta_{calibrated} + 245.94$

where $\theta_{proficiency}$ denotes values on the final transformed scale and $\theta_{calibrated}$ denotes values on the original calibration scale. Overall summary statistics for the trend samples are given in Table 12-8.

As in the past, interpretation of the trend results was facilitated through the provision of scale anchoring information. In 1984, five NAEP reading scale levels were selected as anchor points. These points (described in *Trends in Academic Progress*) are:

150 - Simple, discrete reading tasks;
200 - Partially developed skills and understanding;
250 - Interrelation of ideas and generalizations;
300 - Understanding complicated information; and
350 - Learning from specialized reading materials.

Table 12-8
Means and Standard Deviations on the Reading Trend Proficiency Scale

Age	Assessment	All Five Plausible Values	
		Mean	S. D.
9	1984	211.0	41.1
	1988	211.8	41.2
	1990	209.2	44.7
13	1984	257.1	35.5
	1988	257.5	34.7
	1990	256.8	36.0
17	1984	288.8	40.3
	1988	290.1	37.1
	1990	290.2	41.3

Detailed descriptions of the skills required to read at each level were derived and benchmark exercises were selected to exemplify each level. These same anchor points are used in the 1988 and 1990 reading trend reports. The estimated proportion of students in each reporting category who are at or above each anchor point were examined in *Trends in Academic Progress*.

12.2 CROSS-SECTIONAL DATA ANALYSIS

The data from the main focused-BIB assessment of reading (9[Rdg-MainP], 13[Rdg-MainP], and 17[Rdg-MainP]) and from the answer mode bridge (9[Rdg-ABB]) were used for cross-sectional analyses comparing the levels of reading achievement for various subgroups of the 1990 target populations. The main assessment included three student cohorts: students who were either in the fourth grade or 9 years old, students who were either in the eighth grade or 13 years old, and students who were either in the twelfth grade or 17 years old. The age definition and time of testing for the 9[Rdg-ABB] sample were identical to those used for the 9[Rdg-MainP] sample. The birth date ranges for age-eligible students were based on the 1980, 1976, and 1972 calendar years respectively for ages 9, 13 and 17. The sampled students in each of these three cohorts were assessed either in the winter or the spring. The samples in the main assessment are listed in Table 12-1.

A secondary goal of the analysis of the cross-sectional reading data was, if feasible, to equate the 1990 assessment to the 1988 cross-sectional scale discussed in *Learning to Read in our Nation's Schools: Instruction and Achievement in 1988 at Grades 4, 8, and 12* (Langer, Applebee, Mullis, & Foertsch, 1990) yielding a short-term trend from 1988 to 1990. Forming the link between the 1988 and 1990 assessments was complicated by two factors.

First, for age 9/grade 4 students, there was a change in the response mode format. In 1988, students gridded their responses directly in the test booklet. In 1990, a separate answer sheet was introduced at grade 4 (grade 8 and grade 12 students used separate answer sheets in

both 1988 and 1990). The 9[Rdg-ABB] sample was included to allow the effect of this change to be examined, and to make it possible to equate the 1990 9[Rdg-MainP] to the 1990 9[Rdg-ABB] sample.

A second and more fundamental issue involved the change in the criteria for excluded students. NAEP has always excluded students for whom the assessment would have little meaning, such as students with severe physical or learning disabilities, or student with limited proficiency in English. The 1990 cross-sectional assessment incorporated new rules for the exclusion of students, which were meant to clarify and standardize the criteria for exclusion. If the change in criteria for exclusion was found to have a substantial effect, comparisons between 1988 and 1990 would be of limited value, because the populations being compared would be different. Analyses were conducted to examine the effect of the change in criteria for exclusion and are described in section 12.2.3.

In the cross-sectional samples, each student was administered a booklet containing three blocks of cognitive reading items, a block of background questions common to all booklets for a particular age/grade level, and a block of reading-related background questions common to all reading booklets for a particular age/grade level. Seven blocks of cognitive reading items were administered at each age/grade level in a total of seven booklets for each level. (See Chapter 4 for more information about the blocks and booklets.) Both age- and grade-selected students contributed to the cross-sectional scaling. However, the "grade-only" portion of the main focused-BIB reading samples (whether the time of assessment was winter or spring) contributed to the means and percentages of the cross-sectional results that are reported in *Reading in School and out of School.*

For each grade, approximately one-half of the items in the cross-sectional assessment were identical to items in the 1988 cross-sectional assessment. These items occurred in intact blocks, and provided the common information needed to establish the short-term trend. Table 12-9 gives the blocks and numbers of items common across assessments. .

Table 12-9
1990 Reading Cross-sectional Blocks and Items Common to the 1988 Assessment

Sample	Common Blocks	Number of Common Items
9[Rdg-MainP]	RC, RD, RE	26
9[Rdg-ABB]	RC, RD, RE	26
13[Rdg-MainP]	RC, RD, RE, RF	52
17[Rdg-MainP]	RC, RE, RF	51

The pool of items used in the 1990 reading assessment contained open-ended and multiple-choice questions measuring performance on sets of objectives documented in *Reading Objectives: 1990 Assessment* (NAEP, 1989a). The objectives framework is described in Chapter 2.

A total of 218 distinct reading items addressing these objectives was administered in 1990 using the focused-BIB design to allocate the items to the assessed students.

It should be noted that open-ended items were scattered throughout the seven cognitive blocks for each level. These open-ended items were scored dichotomously and were scaled with the multiple-choice items. In addition, detailed analyses of the ordinal responses to the open-ended items were also conducted, and are summarized in *Reading in School and out of School.* In the final cross-sectional scale, there were 170 multiple-choice items and 5 open-ended items, for a total of 175 items. The answer mode bridge contains an additional 43 items, of which 3 are open-ended. The number of overlapping items for the age/grade levels are listed in Table 12-10. Numbers of items in each subscale by block and by booklet are given in Tables E-1 and E-2 in Appendix E.

Table 12-10
Number of Scaled Reading Items Common Across Grades

Grade	Number of Items
4 [Rdg-ABB] only	43
4 [Rdg-MainP] only	34
8 only	5
12 only	40
4 and 8 only	25
4 and 12 only	5
8 and 12 only	63
4, 8, and 12	3
Total	218

The next sections contain a description of the analysis performed using the main focused-BIB sample data. As in the trend analysis, the process began with an examination of the items and blocks of items. Open-ended items were dichotomously scored and derived background variables were calculated. The effect of changes in criteria for excluding students was then examined, as was the effect of the change in answer mode for age 9/grade 4 students. The estimation of item parameters for the reading scale was completed, and followed by the generation of univariate plausible values. Finally, the plausible values were transformed to the final proficiency scale. Unlike the trend reading scale, the cross-sectional NAEP reading scale was not anchored when it was first administered in 1988. Therefore, the 1990 main reading assessment does not have descriptions of anchor points or exemplar items.

12.2.1 Item Analysis

Tables 12-11, 12-12, 12-13, and 12-14 show the number of items, mean proportion correct, mean r-biserial, and alpha reliability for each block administered for each of the

Table 12-11

Descriptive Statistics for Item Blocks by Position Within Booklet and Over All Occurrences
Reading Answer Booklet Bridge Sample, Grade 4

Statistic	Block Position	Block RC	RD	RE	RJ	RK	RL
Number of scaled items	—	7	7	12	14	13	15
Number of scaled open-ended items	—	0	0	1	0	0	0
Unweighted sample size	1	900	—	908	—	—	900
	2	—	900	899	902	—	—
	3	885	891	—	—	894	—
	ALL	1785	1791	1807	902	894	900
Average weighted proportion correct	1	.72	—	.74	—	—	.73
	2	—	.58	.70	.75	—	—
	3	.65	.54	—	—	.65	—
	ALL	.68	.56	.72	.75	.65	.73
Average weighted r-biserial	1	.89	—	.79	—	—	.71
	2	—	.70	.78	.78	—	—
	3	.88	.70	—	—	.67	—
	ALL	.89	.70	.78	.78	.67	.71
Weighted alpha reliability	1	.78	—	.80	—	—	.80
	2	—	.60	.82	.82	—	—
	3	.81	.61	—	—	.77	—
	ALL	.80	.61	.81	.82	.77	.80
Weighted proportion of students attempting last item	1	.98	—	.92	—	—	.98
	2	—	.99	.89	.96	—	—
	3	.98	.99	—	—	.98	—
	ALL	.98	.99	.91	.96	.98	.98

Table 12-12

Descriptive Statistics for Item Blocks by Position Within Booklet and Over All Occurrences
Reading Cross-sectional Sample, Grade 4

Statistic	Block Position	Block RC	RD	RE	RF	RG	RH	RI
Number of scaled items	——	7	7	12	9	12	11	9
Number of scaled open-ended items	——	0	0	1	1	0	0	1
Unweighted sample size	1	892	912	885	903	890	923	891
	2	898	892	908	883	898	886	922
	3	880	914	891	883	902	874	889
	ALL	2670	2718	2684	2669	2690	2683	2702
Average weighted proportion correct	1	.69	.55	.72	.53	.61	.40	.48
	2	.68	.56	.69	.51	.59	.39	.49
	3	.61	.51	.67	.51	.59	.39	.45
	ALL	.66	.54	.69	.52	.60	.39	.47
Average weighted r-biserial	1	.86	.70	.78	.65	.66	.50	.64
	2	.85	.70	.76	.66	.66	.49	.62
	3	.84	.70	.77	.67	.67	.51	.67
	ALL	.85	.70	.77	.66	.67	.50	.64
Weighted alpha reliability	1	.78	.62	.83	.70	.79	.48	.62
	2	.79	.60	.81	.71	.79	.45	.57
	3	.79	.61	.83	.72	.77	.49	.67
	ALL	.79	.61	.82	.71	.78	.47	.62
Weighted proportion of students attempting last item	1	.97	.98	.85	.85	.90	.90	.87
	2	.96	.99	.87	.86	.90	.92	.88
	3	.99	.99	.87	.86	.94	.91	.89
	ALL	.97	.99	.86	.86	.91	.91	.88

Table 12-13

Descriptive Statistics for Item Blocks by Position Within Booklet and Over All Occurrences
Reading Cross-sectional Sample, Grade 8

Statistic	Block Position	Block						
		RC	RD	RE	RF	RG	RH	RI
Number of scaled items	——	9	10	19	14	15	15	14
Number of scaled open-ended items	——	0	0	0	0	2	0	0
Unweighted sample size	1	946	920	930	931	918	937	922
	2	923	945	920	930	930	917	937
	3	914	935	920	942	917	926	925
	ALL	2783	2800	2770	2803	2765	2780	2784
Average weighted proportion correct	1	.84	.71	.69	.45	.64	.59	.57
	2	.82	.73	.68	.45	.63	.59	.57
	3	.81	.67	.66	.46	.60	.55	.54
	ALL	.82	.70	.68	.45	.62	.57	.56
Average weighted r-biserial	1	.99	.75	.63	.55	.62	.59	.61
	2	.96	.76	.66	.55	.63	.59	.59
	3	.93	.75	.67	.55	.65	.60	.59
	ALL	.96	.75	.65	.55	.63	.60	.59
Weighted alpha reliability	1	.82	.76	.81	.65	.79	.72	.74
	2	.82	.76	.83	.64	.78	.72	.71
	3	.82	.77	.84	.65	.80	.75	.72
	ALL	.82	.77	.82	.65	.79	.73	.72
Weighted proportion of students attempting last item	1	.99	.99	.88	.92	.69	.93	.95
	2	1.00	.98	.87	.93	.73	.93	.95
	3	.99	.98	.90	.93	.72	.95	.95
	ALL	.99	.99	.89	.93	.71	.94	.95

Table 12-14

Descriptive Statistics for Item Blocks by Position Within Booklet and Over All Occurrences
Reading Cross-sectional Sample, Grade 12

Statistic	Block Position	Block						
		RC	RD	RE	RF	RG	RH	RI
Number of scaled items	——	18	15	19	14	14	17	14
Number of scaled open-ended items	——	0	1	0	0	1	0	1
Unweighted sample size	1	895	886	890	901	902	872	899
	2	898	895	887	889	901	901	871
	3	901	869	898	894	886	888	899
	ALL	2694	2650	2675	2684	2689	2661	2669
Average weighted proportion correct	1	.77	.72	.80	.62	.72	.65	.69
	2	.76	.74	.80	.63	.73	.64	.69
	3	.74	.70	.78	.62	.71	.64	.68
	ALL	.76	.72	.79	.63	.72	.64	.69
Average weighted r-biserial	1	.75	.67	.70	.64	.64	.62	.67
	2	.76	.67	.73	.63	.65	.65	.66
	3	.76	.71	.78	.66	.68	.64	.69
	ALL	.76	.69	.74	.64	.66	.63	.68
Weighted alpha reliability	1	.78	.80	.81	.72	.75	.79	.77
	2	.82	.77	.83	.70	.76	.81	.76
	3	.82	.81	.85	.74	.77	.81	.78
	ALL	.81	.80	.83	.72	.76	.80	.77
Weighted proportion of students attempting last item	1	.93	.78	.89	.95	.83	.83	.76
	2	.90	.79	.88	.94	.84	.85	.75
	3	.93	.85	.92	.95	.88	.84	.79
	ALL	.92	.81	.90	.94	.85	.84	.77

samples. These values were calculated for the dichotomously scored multiple-choice and open-ended items within a block, if they were used in the scaling process. The table also gives the number of students who were administered the block and the percent not reaching the last item in the block. These numbers include only those students in the grade-only portion of the samples that contributed to the summary statistics provided in *Reading in School and out of School*. Student weights were used for all statistics, except for the sample sizes. The results for the blocks administered to each age/grade level indicated that the blocks differ in number of items, average difficulty, reliability, and percent not reaching the last item, and so are not parallel to one another. Preliminary item analyses for all items within a block were completed before scaling; however, the results shown here indicate the characteristics of the items that contributed to the final scale.

Tables 12-11 through 12-14 also contain information about the effect of the position of blocks within booklets on the average percent correct for items within each block presented to the focused-BIB samples for each grade. The order of blocks within booklets did have an effect: In most cases, the average percent correct declined as the position moved from 1 to 3.

12.2.2 Scoring the Open-ended Items

As indicated earlier, the reading assessment included open-ended items. Responses to these items were dichotomized and included in the scaling process. In addition, detailed analyses of the ordinal responses to the open-ended items were also conducted, and are summarized in *Reading in School and out of School*. Chapter 7 contains reliability information (the means and ranges for percent agreement between raters) for the items as they were originally scored. The right/wrong scoring of the categories of responses for the items are indicated in Table G-1 in Appendix G. The percent agreement for the raters and Cohen's Kappa, calculated after the items were dichotomized, are also shown in the table. The sample sizes listed in this table represent the size of the samples used in calculating the rater reliability information; that is, the number of responses that were rescored for each item.

Only one open-ended reading item (R000806) was common to both the 1988 and 1990 cross-sectional assessments. The professional scoring of this item was based on a different rubric in 1990 than in 1988; even the number of response categories differed between the two assessments. Therefore, this item was treated as if it were two separate items for the purposes of the short-term trend.

12.2.3 Excluded Students

As was noted above, the 1990 assessment saw the introduction of new, more specific criteria for the exclusion of students from the NAEP assessment (see Chapter 5). If the change in the definition for exclusion were to have a substantial effect, the utility of the short-term trend scale would be called into question, because changes in proficiency between the 1988 and 1990 samples would be confounded with changes in the characteristics of the population being assessed. As a hypothetical example, suppose that the 1988 rules for exclusion resulted in most students with dyslexia being excluded from the assessment, while the 1990 rules resulted in students with all but the most severe disabilities being assessed. If the mean for the 1990

assessment were lower than that for the 1988 assessment, it would be impossible to say whether the result is due to lower reading proficiency for average students, or due to the fact that students with dyslexia were included in 1990 but not in 1988.

Therefore, it was necessary to assess the effect of the change in criteria for exclusion from the assessment. The simplest method is to compare rates of exclusion of students under the new criteria to those obtained under the old criteria for exclusion. The 1990 trend (bridge) samples were drawn using the old criteria for exclusion,[3] while the 1990 main samples were obtained using the new criteria. To be strictly comparable, however, the samples must be assessed at the same time of year and have identical age definitions. As shown in Table 12-1, the age 9/grade 4 main sample, which was assessed in the winter, is comparable to the bridge samples in terms of time of testing. Similarly, the age 17/grade 12 sample and the age 17 bridge sample were both assessed in the spring. However, the age definitions (calendar year and not-calendar year) differ; the only overlap is for students born between October and December of 1972. There is no overlap for the age 13/grade 8 sample. Comparisons of the other samples, while not as precise, may still be informative. Table 12-15 gives the rate of exclusion for various NAEP samples. Shaded entries indicate samples within a column that are strictly comparable.

Table 12-15
Percentage of Students Excluded

Sample	Age 9/Grade 4	Age 13/Grade 8	Age 17 (born 10/72-12/72)
1990 Trend (Bridge Samples)	6.6	6.4	4.2
1990 Cross-Sectional (Winter)	6.8	6.6	4.5
1990 Cross-Sectional (Spring)	5.8	5.9	3.1
1990 Main (Overall)	6.3	6.2	3.8
1988 Main (Overall)	4.9	6.2	3.2*

* This entry is for the full age 17/grade 12 sample.

As shown in Table 12-15, while there were some differences in the pattern of exclusion, these were small, and the overall pattern of exclusion was similar for both sets of criteria. However, the fact that the rates of exclusion were similar does not guarantee that the same students are excluded under the two sets of criteria. Therefore, additional analyses were performed to compare selected background characteristics of students who were excluded from the assessment. Based upon these analyses (summarized in Table 12-16), it was concluded that, while some differences existed, they were not of a magnitude to render short-term trend

[3] For the purposes of exclusion or inclusion of students in the sample, no distinction was made between the bridge-to-1984 and the bridge-to-1986 samples. Neither was any distinction made between the subject areas. Each of these assignments was made subsequent to the decision whether to include a given student in the assessment.

Table 12-16

Percentages and Jackknifed Standard Errors for Subgroup Memberships of Students Excluded from the 1990 Assessment

Age/Grade	Subgroup		Sample: Bridge	Sample: Main Winter	Sample: Main Spring
Age 9/Grade 4	Gender:	Male	67.4 (1.9)	65.2 (1.3)	67.0 (2.2)
		Female	32.6 (1.9)	34.7 (1.4)	32.5 (2.2)
		Omit	0.0 (0.0)	0.1 (0.1)	0.5 (0.4)
	Race/ethnicity:	White	53.7 (3.6)	55.5 (3.8)	50.9 (3.1)
		Black	15.8 (2.5)	16.8 (1.7)	20.8 (2.0)
		Hispanic	16.2 (2.3)	15.2 (2.0)	15.1 (2.1)
		Asian American	3.8 (0.9)	2.1 (0.5)	3.8 (1.5)
		American Indian/Alaska Native	0.3 (0.2)	0.6 (0.3)	0.2 (0.2)
		Other	1.2 (2.5)	0.4 (0.1)	0.7 (0.3)
		Omit	9.1 (2.5)	9.4 (2.8)	8.5 (2.0)
	Reason for Exclusion:	Disabled/IEP	54.2 (3.5)	62.0 (3.7)	64.4 (3.9)
		Limited English Proficiency	19.4 (2.5)	14.7 (1.9)	16.5 (2.9)
		Both	1.6 (0.5)	1.6 (0.4)	1.7 (0.6)
		Nonreader	3.9 (1.3)	3.8 (1.2)	1.7 (0.6)
		Other	8.6 (1.7)	9.1 (2.8)	5.4 (2.6)
		Multiple Response	0.1 (0.1)	0.0 (0.0)	0.1 (0.1)
		Omit	12.2 (3.2)	8.8 (2.2)	10.2 (2.0)
Age 13/Grade 8	Gender:	Male	63.2 (2.1)	66.0 (1.6)	64.7 (1.7)
		Female	36.7 (2.1)	33.8 (1.6)	35.3 (1.7)
		Omit	0.1 (0.1)	0.1 (0.1)	0.0 (0.0)
	Race/ethnicity:	White	49.3 (2.7)	49.9 (3.1)	49.0 (3.4)
		Black	19.7 (2.6)	19.0 (2.5)	19.5 (2.5)
		Hispanic	20.5 (2.9)	18.1 (2.3)	19.0 (2.1)
		Asian American	3.2 (1.0)	3.4 (1.4)	2.4 (0.8)
		American Indian/Alaska Native	0.0 (0.0)	0.8 (0.3)	0.3 (0.2)
		Other	0.5 (0.3)	0.3 (0.2)	0.9 (0.5)
		Omit	6.8 (2.1)	8.3 (2.6)	8.9 (1.4)
	Reason for Exclusion:	Disabled/IEP	66.6 (4.5)	71.9 (3.3)	70.6 (3.8)
		Limited English Proficiency	16.9 (3.3)	14.6 (2.5)	15.1 (3.2)
		Both	1.4 (0.3)	2.6 (0.8)	1.0 (0.4)
		Nonreader	0.4 (0.3)	0.6 (0.3)	0.0 (0.0)
		Other	5.8 (1.4)	3.6 (1.5)	2.8 (1.0)
		Multiple Response	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
		Omit	8.9 (2.1)	6.7 (2.0)	10.6 (1.9)
Age 17/Grade 12	Gender:	Male	65.3 (2.9)	72.5 (4.2)	65.4 (4.5)
		Female	34.3 (2.9)	27.5 (4.2)	34.2 (4.5)
		Omit	0.5 (0.2)	0.0 (0.0)	0.4 (0.4)
	Race/ethnicity:	White	53.0 (3.8)	50.8 (5.8)	49.0 (5.7)
		Black	14.5 (2.8)	21.1 (5.0)	17.9 (4.4)
		Hispanic	18.6 (2.9)	15.8 (3.5)	19.6 (4.2)
		Asian American	5.8 (1.6)	4.7 (2.4)	3.9 (2.1)
		American Indian/Alaska Native	0.3 (0.3)	0.0 (0.0)	0.0 (0.0)
		Other	0.7 (0.6)	0.6 (0.6)	0.0 (0.0)
		Omit	7.1 (1.8)	7.1 (3.6)	9.7 (3.3)

Table 12-16 (continued)

Percentages and Jackknifed Standard Errors for Subgroup Memberships of Students Excluded from the 1990 Assessment

Age/Grade	Subgroup		Sample: Bridge	Sample: Main Winter	Sample: Main Spring
Age 17/Grade 12	Reason for Exclusion:	Disabled/IEP	59.2 (4.5)	66.9 (6.7)	58.6 (6.3)
		Limited English Proficiency	18.7 (2.4)	15.1 (4.0)	18.2 (5.7)
		Both	1.9 (1.1)	0.0 (0.0)	0.0 (0.0)
		Nonreader	1.6 (1.2)	0.9 (1.0)	0.0 (0.0)
		Other	5.1 (2.3)	11.0 (4.9)	5.4 (2.4)
		Multiple Response	1.0 (0.7)	0.0 (0.0)	1.2 (0.8)
		Omit	12.5 (2.8)	6.0 (2.8)	16.5 (5.3)

comparisons meaningless. Thus, it was determined to proceed with the formation of the short-term trend scale.

12.2.4 Effect of Answer Mode

A second difficulty in the formation of the short-term trend scale was the change in answer mode for the age 9/grade 4 sample. In the 1988 assessment, these students gridded their responses directly into their test booklets. In 1990, age 9/grade 4 students responded on a separate answer sheet.[4] The 9[Main-ABB] sample was included in the 1990 cross-sectional assessment to examine any effect this change might have had, and to provide a bridge to equate the 9[Rdg-MainP] sample to the 1988 scale.

To examine the effect of changing the answer mode, a subset of the 9[Rdg-MainP] sample was compared to the answer mode bridge sample. To achieve maximal comparability, the comparison was restricted to items occurring in intact blocks that appeared in the identical positions within the booklet. Some results of this comparison are summarized in Table 12-17.

Table 12-17
Mean Percent Correct by Answer Mode
Reading Cross-Sectional Assessment, Age 9/Grade 4

Group	Answer Mode		Difference
	In Booklet	Separate Answer Sheet	
Male	59.3	56.2	3.1
Female	63.8	62.7	1.1
White	65.8	63.8	2.0
Black	49.0	47.4	1.7
Hispanic	51.8	47.7	4.1
Total	61.5	59.4	2.2

Comparison of the answer modes revealed a moderate, consistent effect. Items were harder when the students were required to respond on a separate answer sheet. Hence, an additional equating step was necessary to place the 9[Rdg-MainP] responses on the 1988 scale. A five-step process was used to place all of the 1990 grade 4 data on the 1988 scale:

1) The items from the 9[Rdg-ABB] sample were calibrated with data from the 1988 assessment to form a short-term trend. (This process is described in the next section.) Using these results in combination with examinee background information, plausible values were generated for the 9[Rdg-ABB] sample.

[4] Age 13/grade 8 and age 17/grade 12 students responded on separate answer sheets in both years.

2) The items from the 9[Rdg-Main] sample were calibrated separately. Using these results in combination with examinee background information, plausible values were generated for the 9[Rdg-Main] sample.

3) Because students in the 9[Rdg-ABB] sample and the 9[Rdg-Main] sample are two random samples of students from the same population, a linear common population equating was used to transform the plausible values for the 1990 main sample onto the arbitrary scale derived from the calibration of 1990 answer mode bridge data.

4) Plausible values for the 9[Rdg-ABB] sample were transformed to place them on the 1988 (short-term trend) scale.

5) The same transformation was then applied to the transformed 1990 main scores obtained in step 3. This placed the 9[Rdg-Main] plausible values on the short-term trend scale.

12.2.5 Estimation of Item Parameters

The first step in the scaling process was the estimation of item parameters for the trend items. Items were calibrated separately for each of the three age groups using the NAEP-BILOG program. Item parameters were estimated using combined data from the assessment years 1988 and 1990, treating each assessment as a sample from a separate subpopulation. The calibration was performed using all the available examinees. Student sampling weights were used for the analysis.

Based upon item analysis, one item (R011601) was deleted from scaling, due to a low biserial correlation with the rest of the items in the block. Further examination revealed that this item had a nonmonotonic empirical IRF. Empirical IRFs for each assessment were compared to the common IRF fit by NAEP-BILOG. No additional items were deleted from the IRT analysis or treated as different items for separate groups due to lack of fit to the model.

Three separate scalings, one for each age/grade sample, were conducted. Because the answer mode was found to have an effect at age 9/grade 4, a fourth analysis was necessary. The analyses were conducted on the following samples:

- the 1988 age 9/grade 4 sample with the 1990 age 9/grade 4 answer mode bridge sample;
- the 1990 age 9/grade 4 main cross-sectional sample;
- the 1988 age 13/grade 8 sample with the 1990 age 13/grade 8 main cross-sectional sample; and
- the 1988 age 17/grade 12 sample with the 1990 age 17/grade 12 main cross-sectional sample.

For each calibration involving multiple samples, a preliminary scaling was performed using only items that were administered in both samples. A second calibration was then performed separately for each assessment. This calibration held constant the parameters of items used in the first calibration while estimating parameters for those items calibrated separately across assessments (including open-ended items). A list of the items scaled for each of the samples, along with their item parameter estimates, appears in Tables E-11 through E-14 in Appendix E.

12.2.6 Generation of Plausible Values

Univariate plausible values were generated for each sample (each age/grade group and the answer mode bridge sample) separately using the NAEP-MGROUP program. Final student weights were used in this analysis. The codings of the reading-specific conditioning variables are presented in Appendix F in Table F-2. Common core conditioning variables were also used. The codings of the common conditioning variables are given in Appendix F in Table F-1. (For age 17/grade 12, the "modal age, > modal grade" category was deleted from the age-by-grade variable, because students above grade 12 were not sampled.) The estimated conditioning effects for the background contrasts of the four samples (defined by the three age/grade groups and the grade 4 answer mode bridge sample) are given in Appendix F in Tables F-8 through F-11. The values of the conditioning effects are expressed in the metrics of their original calibration scales.

12.2.7 The Final Proficiency Scale

The linear indeterminacy of the trend scale was resolved by linking the 1990 short-term trend scale to the 1988 cross-sectional scale. For grades 8 and 12, the item parameters from 1990 based on data from 1988 and 1990 were used with the 1988 data to find plausible values for the 1988 data. The mean and standard deviation of all of the plausible values was calculated and matched to the mean and standard deviation of all of the plausible values based on the 1988 item parameters and 1988 data as reported in earlier reports. The transformations that resulted from this matching of the first two moments for the 1988 data are

Grade 8: $\theta_{\text{proficiency}} = 39.33 \bullet \theta_{\text{calibrated}} + 255.92$

Grade 12: $\theta_{\text{proficiency}} = 36.87 \bullet \theta_{\text{calibrated}} + 283.55$

where $\theta_{\text{proficiency}}$ denotes values on the final transformed scale and $\theta_{\text{calibrated}}$ denotes values on the original calibration scale.

The change in answer mode necessitated an additional transformation at grade 4. First, the 9[Rdg-MainP] sample was transformed to the answer mode bridge scale:

Grade 4: $\theta_{\text{ABB}} = 0.9997 \bullet \theta_{\text{Main}} + 0.0112$.

Next, the 9[Rdg-ABB] scores were transformed to the 1988 scale. The same transformation was applied to the transformed scores for the 9[Rdg-MainP] sample, placing all of the grade 4 data on the short-term trend scale:

$$\text{Grade 4: } \theta_{proficiency} = 43.82 \bullet \theta_{ABB} + 225.45 \ .$$

The 9[Rdg-ABB] and 9[Rdg-Main] samples were then combined. Overall summary statistics for the short-term trend samples are given in Table 12-18.

Table 12-18
Means and Standard Deviations on the Reading Short-term Trend Proficiency Scale

Grade	Assessment	All Five Plausible Values	
		Mean	S. D.
4	1988 1990	230.4 233.0	41.4 42.0
8	1988 1990	262.8 260.6	37.3 37.4
12	1988 1990	287.1 288.5	34.8 35.0

12.2.8 Partitioning of the Estimation Error Variance

For each grade, the variance of the final, transformed scale mean was partitioned as described in Chapter 11. This analysis yielded estimates of the proportion of error variance due to sampling students and the proportion due to the latent nature of θ. Table 12.19 contains estimates of the sampling variance U^* and a multiple of the estimates of the variance among the weighted means of the five plausible values B. The table also contains an estimate of the total error variance, V, as well as the proportion of error variance due to sampling students and due to the latent nature of θ.

Table 12-19
Estimation Error Variance and Related Coefficients for the Reading Cross-sectional Assessment

Grade	U^*	$(1+5^{-1})B$	V	Proportion of Variance Due to...	
				Student Sampling: U^*/V	Latency of θ: $(1+5^{-1})B/V$
4	0.72	0.05	0.76	0.94	0.06
8	0.93	0.02	0.95	0.98	0.02
12	0.98	0.03	1.01	0.97	0.03

12.2.9 Mantel-Haenszel DIF Analyses

Differential item functioning (DIF) analyses of the main-BIB reading items used in scaling were completed after scaling to provide information to test developers for developing future reading assessments. Sample sizes were large enough to compare male and female students, White and Black students, and White and Hispanic students using the Mantel-Haenszel procedure described in Chapter 9 to identify items that should be examined more closely ("C" items). DIF analyses were conducted separately by grade. A given item was subjected to at least three, and as many as nine, separate DIF analyses. Table 12-20 summarizes information about the identified items for each block.

Ten of the 218 cross-sectional reading items were categorized as "C" items in one or more of the DIF analyses. Table H-1 (in Appendix H) identifies items that were categorized as "C" items in at least one analysis. The block containing the item, the grade, and the analysis for which the item was identified are also presented in Table H-1. Only two items, N001601 and R011803, were found to have possible bias by a committee of test and subject-area experts convened to examine the "C" items. Both items involved vocabulary questions from the reading passage that the committee felt would be difficult to infer from context, and that were likely to be differentially familiar to one of the groups. Item N001601 showed DIF against Hispanic students; item R011803 showed DIF against female students. Neither of these items will be retained for future reading assessments.

Table 12-20
Numbers of "C" Items Favoring Each Group by Reading Block

Block	Grade	Number of Scaled Items in Block	Analysis					
			Male/Female		White/Black		White/Hispanic	
			Male	Female	White	Black	White	Hispanic
RC	4	7	0	0	0	0	0	0
	8	9	0	0	0	0	0	0
	12	18	0	0	0	0	0	0
RD	4	7	0	0	0	0	0	0
	8	10	0	0	0	0	1	0
	12	15	0	1	0	0	0	0
RE	4	12	1	0	1	0	0	0
	8	19	0	1	0	2	0	0
	12	19	0	1	1	1	0	0
RF	4	9	0	0	0	0	0	0
	8	14	0	0	0	0	0	0
	12	14	0	0	1	0	0	0
RG	4	12	0	0	1	0	0	0
	8	15	0	0	0	0	0	0
	12	14	0	0	1	0	0	0
RH	4	11	0	0	0	0	0	0
	8	15	0	0	0	0	0	0
	12	17	0	0	0	0	0	0
RI	4	9	0	0	0	0	0	0
	8	14	0	0	0	0	0	0
	12	14	1	0	0	0	0	0

Chapter 13

DATA ANALYSIS FOR THE MATHEMATICS ASSESSMENT[1]

Kentaro Yamamoto and Frank Jenkins

Educational Testing Service

This chapter describes the analyses performed on the responses to the cognitive and background items in the 1990 assessment of mathematics. These analyses led to the results presented in *Trends in Academic Progress: Achievement of U.S. Students in Science, 1969-0 to 1990; Mathematics 1973 to 1990; Reading, 1971 to 1990; and Writing, 1984 to 1990* (Mullis, Dossey, Foertsch, Jones, & Gentile, 1991) and in *The State of Mathematics Achievement: NAEP's 1990 Assessment of the Nation and the Trial Assessment of the States* (Mullis, Dossey, Owen, & Phillips, 1991). The emphasis of this chapter is on the methods and results of procedures used to develop the IRT-based scale scores that formed the basis of these reports. However, some attention is given to the analysis of open-ended items also reported in *The State of Mathematics Achievement*. The theoretic underpinnings of the IRT and plausible values methodology described in this chapter are given in Chapter 11.

The objectives of the mathematics analyses were to

- prepare scale values and perform all analyses necessary to produce a long-term trend report in mathematics. The mathematics trend line includes the years 1973, 1978, 1982, 1986, and 1990.

- prepare scale values for the cross-sectional analysis of the main focused-BIB mathematics samples. The scaling of mathematics entailed development of several subscales and an overall composite.

The student samples that were administered mathematics items in the 1990 assessment are shown in Table 13-1. (See Chapters 1 and 3 for descriptions of the target populations and the sample design used for the assessment.) Data from the first six samples (Math-MainP and Math-MainT) were used in the cross-sectional analysis, while data from the rest of the samples were collected for trend purposes.

The cross-sectional data from the 1990 main focused-BIB samples were scaled separately from the data of the trend (bridge) samples. Accordingly, the trend and cross-sectional analyses are presented in separate sections. Section 13.1 pertains to the scaling of the data from the

[1]Data analysis and scaling were performed by Edward Kulick, Drew Bowker, Wing Lowe, and Steven Isham. Nancy Allen assisted with part of the analysis.

trend bridges; section 13.2 contains information about the scaling of the data from the main focused-BIB samples.

Table 13-1
NAEP 1990 Mathematics Student Samples

Sample	Booklets	Mode	Cohort Assessed	Time of Testing	Age Definition	Modal Grade	Number Assessed
9 [Math-MainP]	11-17	Print	Age 9/grade 4	Winter, spring	CY	4	8,790
13 [Math-MainP]	8-14	Print	Age 13/grade 8	Winter, spring	CY	8	8,634
17 [Math-MainP]	8-14	Print	Age 17/grade 12	Winter, spring	CY	12	8,406
9 [Math-MainT]	28	Tape	Age 9/grade 4	Winter, spring	CY	4	3,187
13 [Math-MainT]	25	Tape	Age 13/grade 8	Winter, spring	CY	8	3,182
17 [Math-MainT]	25	Tape	Age 17/grade 12	Winter, spring	CY	12	3,139
9 [RMS-Br86]	91-93	Mixed	Age 9	Winter	CY	4	6,235
13 [RMS-Br86]	91-93	Mixed	Age 13	Fall	CY	8	6,649
17 [RMS-Br86]	61-66	Print	Age 17/grade 11	Spring	Not CY	11	8,338
17 [MS-Br86]	84-85	Tape	Age 17	Spring	Not CY	11	4,411
9 [MS-BrLT]	94-95	Tape	Age 9	Winter	CY	4	4,134
13 [MS-BrLT]	94-95	Tape	Age 13	Fall	CY	8	4,455
17 [MS-BrLT]	94-95	Tape	Age 17	Spring	Not CY	11	4,402

LEGEND:

Math Mathematics
RMS Reading, mathematics, and science
MS Mathematics and science

MainP Main assessment, print administration
MainT Main assessment, tape administration

Br86 Bridge to 1986
BrLT Bridge for long-term trend

Print Printed administration
Tape Audiotape administration
Mixed Mathematics and science administered by audiotape, reading administered by print

CY Calendar year: birthdates in 1980, 1976, and 1972 for ages 9, 13, and 17

Not CY Age 17 only: birthdates between Oct. 1, 1972 and Sept. 30, 1973

13.1 TREND DATA ANALYSIS

The trend results reported in *Trends in Academic Progress* are based on paced-tape administrations and occur at all of the age levels. The samples involved in the analysis are shown as 9[RMS-Br86], 13[RMS-Br86], and 17[MS-Br86] in Table 13-1. For ages 9 and 13, the trend booklets for these samples contained blocks of reading, mathematics, and science items. The science and mathematics blocks were paced by tape recordings and the reading blocks were presented in print form. The age 17 trend booklets contained only mathematics and science blocks, both presented by paced-tape recordings. All students received a block of common background questions, distinct for each age. Subject-area background questions were presented in the cognitive blocks. The booklets for the age 9 and age 13 samples (booklets 91-93) are the same as those used for trend assessments in 1986 and 1988. The booklets for the age 17 sample (booklets 84-85) are the same as those used for the 1986 trend assessment. The booklets and

the blocks within those booklets are listed in Tables 4-7, 4-8, and 4-9 of Chapter 4. Additional information about all of the items in these blocks can be found in Tables 4-10, 4-11, and 4-12 of that chapter. This chapter includes specific information about the trend items that were scaled.

Table 13-2 clarifies the relationships between the 1990 trend samples and samples from previous years. For ages 9 and 13, the paced-tape bridge to the 1986 samples allow direct comparisons with 1988 samples (a year in which the results were published only in *Disentangling the NAEP 1985-86 Reading Anomaly* [Beaton & Zwick, 1990]), as well as with 1986 trend samples. For age 17, the paced-tape sample (17[MS-Br86]) does not allow direct comparisons with a 1988 sample, but comparisons can be made with 1986 trend samples. In 1986, the mathematics trend items were scaled with common items from the 1977-78 and 1981-82 assessments. Because the 1972-73 assessment had few items in common with the current assessment, data from that assessment was not scaled using the IRT model but was linked to the trend line by a linear transformation involving the mean proportion correct for common items. The 1990 trend assessment was linked to the 1972-73, 1977-78, and 1981-82 assessments through the 1986 assessment. Information about previous assessment years is available in *Expanding the New Design: The NAEP 1985-86 Technical Report* (Beaton, 1988).

In addition to the samples that contributed data to the results reported in *Trends in Academic Progress*, two other types of trend samples were collected in 1990. One is a print sample that is comparable to an assessment of age 17 students that was part of the 1988 study of the 1986 reading anomaly. The print trend sample, 17[RMS-Br86], provides the link to the 1988 trend data. The booklets, presentation, time of assessment, and age definition were the same as those for the 1988 mathematics trend sample for which results have been published in *Disentangling the NAEP 1985-86 Reading Anomaly* (Beaton & Zwick, 1990). Because the results from the age 17 print trend sample were not included in *Trends in Academic Progress*, analyses of these data are not described here.

The other assessment not contributing to the results in *Trends in Academic Progress* consisted of new booklets formed with items from previous assessments. The samples that were assessed using these booklets include 9[MS-BrLT], 13[MS-BrLT], and 17[MS-BrLT]. For these samples we used the same time of administration, mode, and age definitions as the samples that bridge to 1986. The purpose of these samples was to bolster the number of items and students contributing to long-term trend. Future trend assessments were to include the items presented to these samples. The items in these new booklets were not previously scaled. The reasons that the data from these samples were not included in *Trends in Academic Progress* are detailed in section 13.1.2.

The steps in the mathematics trend analysis are documented in the following sections. As is usual in NAEP analyses, the first step was to gather item and block information. Next, the trend items were calibrated and plausible values were generated after conditioning on available background variables. Derived background variables were calculated. Finally, the scales were placed on the final mathematics trend proficiency scale used in previous trend assessments.

Table 13-3 indicates the number of items common across different age combinations. Table 13-4 shows the number of items (scaled in 1990) that were common across assessment years. The 1986, 1988, and 1990 assessments had all items in common. For age 9 the number

Table 13-2

NAEP Mathematics Samples Contributing to 1990 Trend Results, 1973-1990

Cohort	Year	Sample	Subjects	Time of Testing	Mode of Administration	Age Definition	Modal Grade
Age 9	1973	MainA	SM	Winter	Tape	CY	4
	1978	MainB	M	Winter	Tape	CY	4
	1982	MainC	MCS	Winter	Tape	CY	4
	1986	BrLT*	RMS	Winter	Mixed	CY	4
	1988	Br86*	RMS	Winter	Mixed	CY	4
	1990	Br86*	RMS	Winter	Mixed	CY	4
Age 13	1973	MainA	MS	Fall	Tape	CY	8
	1978	MainB	M	Fall	Tape	CY	8
	1982	MainC	MCS	Fall	Tape	CY	8
	1986	BrLT*	RMS	Fall	Mixed	CY	8
	1988	Br86*	RMS	Fall	Mixed	CY	8
	1990	Br86*	RMS	Fall	Mixed	CY	8
Age 17	1973	MainA	SM	Spring	Tape	Not CY	11
	1978	MainB	M	Spring	Tape	Not CY	11
	1982	MainC	MCS	Spring	Tape	Not CY	11
	1986	BrLT*	MS	Spring	Tape	Not CY	11
	1990	Br86*	MS	Spring	Tape	Not CY	11

* Within an age group these samples had common booklets and constituted a trend line.

LEGEND:

M	Mathematics
MCS	Mathematics, civics, and science
MS	Mathematics and science
RMS	Mathematics, reading, and science
SM	Science and mathematics
Print	Printed administration
Tape	Audiotape administration
Mixed	Mathematics and science administered by audiotape, reading administered by print
MainA	Main assessment 1973
MainB	Main assessment 1978
MainC	Main assessment 1982
Br86	Bridge to 1986 (trend)
BrLT	Long-term bridge (trend)
CY	Calendar year: birthdates in 1980, 1976, and 1972 for ages 9, 13, and 17
Not CY	Age 17 only: birthdates between Oct. 1, 1972 and Sept. 30, 1973

Table 13-3

Numbers of Scaled Mathematics Trend Items Common Across Ages

Age	Booklets	Number of Items
9 only	91-93	32
13 only	91-93	30
17 only	84-85	41
9 and 13 only	91-93, 91-93	20
9 and 17 only	91-93, 84-85	0
13 and 17 only	91-93, 84-85	27
9, 13, and 17	91-93, 91-93, 84-85	3
Total		153

Table 13-4

Numbers of Scaled Mathematics Trend Items Common Across Assessments

Assessment Year	Number of Items
Age 9	
1986, 1988, 1990	55
1982, 1986, 1988, 1990	53
1978, 1986, 1988, 1990	35
1978, 1982, 1986, 1988, 1990	35
Age 13	
1986, 1988, 1990	80
1982, 1986, 1988, 1990	79
1978, 1986, 1988, 1990	56
1978, 1982, 1986, 1988, 1990	56
Age 17	
1986, 1990	71
1982, 1986, 1990	65
1978, 1986, 1990	54
1978, 1982, 1986, 1990	54

of items common across assessment years 1978 to 1990 was only 35. For age 13 the overlap across all assessments was 56 items and for age 17 the overlap was 54 items. Item parameters were estimated assuming a univariate scale since the number of items presented to each student was small and there were too few items to estimate several subscales separately.

In the first phase of the analysis, standard item statistics were calculated. The results serve as a check for data entry errors and as a reasonableness check against results from previous assessments.

In the second phase, the IRT model was fitted to the data across multiple assessments for each age separately. This puts item parameters and ability estimates on the same scale across years. Note that the same item may have different item parameters for different age groups.

The analysis for an age group was completed by the creation of plausible values through a multiple imputation estimation procedure in which item parameter estimates, student responses and student background information are combined to produce the most precise possible estimates of student ability. Plausible values from the 1990 assessment were transformed to the scale of the 1986 proficiency measures.

The 1990 plausible values for ages 9, 13, and 17 were used to create proficiency means and jackknifed standard errors for the whole group and for subgroups. These proficiency means form the final point in the longitudinal mathematics trend from 1973 to 1990.

The specifics of the mathematics trend analysis are documented in the following sections.

13.1.1 Item Analysis for the Paced-tape Bridge-to-1986 Trend Assessment

The conventional item analysis, with results displayed in Table 13-7, was performed at the block level on the paced tape trend data. No problems in coding, formats, or data were detected. The correspondence between blocks, booklets, and samples is given for the regular trend assessment in Table 13-5 and for the supplemental trend assessment in Table 13-6. Note that common labeling of these blocks across ages does not denote common items.

Table 13-7 contains the number of items, size of the sample administered the block, mean weighted proportion correct, mean weighted r-biserial, and mean weighted alpha as a measure of reliability for each block. The average values were calculated using examinee weights and the items in the block that were scaled. The 1990 item-level statistics were not very different from those for the 1986 and 1988 assessments. The percent of examinees not reaching items in the trend blocks was almost always zero because the items were administered with a tape recording to pace response time.

13.1.2 Item Analysis for the Paced-tape Trend Supplement

In order to augment the item pool and the sample size for the 1990 paced tape bridge-to-1986 trend assessment, a supplemental trend sample was selected at all three ages (samples

Table 13-5

Correspondence Between Samples, Booklets, and Blocks for Bridge-to-1986 Mathematics Trend

Sample	Booklet	Blocks
Age 9 RMS-Br86	91	M1
	92	M3
	93	M2
Age 13 RMS-Br86	91	M1
	92	M3
	93	M2
Age 17 MS-Br86	84	M1 M2
	85	M3

Table 13-6

Correspondence Between Samples, Booklets, and Blocks for Supplemental Mathematics Trend

Sample	Booklet	Blocks
Age 9 MS-BrLT	94	MM MN
	95	MO
Age 13 MS-BrLT	94	MM MN
	95	MO
Age 17 MS-BrLT	94	MM MN
	95	MO

Table 13-7

Descriptive Statistics for Item Blocks
Mathematics Trend Samples

Statistic	Block M1	Block M2	Block M3*
Age 9			
Number of scaled items	24	26	5
Number of scaled open-ended items	9	9	0
Unweighted sample size	1991	2194	2050
Average weighted proportion correct	.63	.62	.68
Average weighted r-biserial	.59	.63	.82
Weighted alpha reliability	.81	.86	.46
Age 13			
Number of scaled items	36	36	8
Number of scaled open-ended items	9	8	0
Unweighted sample size	2229	2132	2288
Average weighted proportion correct	.66	.60	.65
Average weighted r-biserial	.56	.57	.66
Weighted alpha reliability	.85	.86	.58
Age 17			
Number of scaled items	33	33	5
Number of scaled open-ended items	10	5	1
Unweighted sample size	2205	2205	2206
Average weighted proportion correct	.64	.65	.54
Average weighted r-biserial	.69	.64	.75
Weighted alpha reliability	.91	.88	.54

* This block is mostly calculator items, which were not analyzed.

9[MS-BrLT], 13[MS-BrLT], and 17[MS-BrLT]). These samples were administered supplemental blocks of trend items not included in the regular trend blocks. In 1986 the regular trend blocks (M1, M2, and M3) were composed of the best available items from previous assessments. The supplemental trend blocks (MM, MN, and MO) were composed of the best available items not already in the regular trend blocks. Table 13-6 gives the correspondence between trend supplement samples, booklets, and blocks.

The original analysis plan was to scale the items from the regular and supplemental trend blocks together, so it was important that the two sets of items be drawn from the same universe; that is, that the two sets of items not exhibit greatly different measurement properties. As it turned out, the item analysis indicated that there were important differences in the results between the regular and the supplemental trend samples. Since the two samples were drawn from the same population, these differences were presumed to be due to the different characteristics of the regular and supplemental trend item pools. Table 13-8 presents the subgroup mean percent correct values of males and females receiving the regular blocks and the supplemental blocks. For ages 13 and 17, the regular trend blocks evidenced a difference between male and female means that was somewhat larger and more in favor of males than the supplemental trend blocks. The gender differences displayed by the regular trend blocks were congruent with those found in the 1986 trend analysis. As a result, it was decided that the supplemental trend blocks altered the nature of the assessment by altering somewhat the observed performance of gender subgroups. Discrepancies in gender differences between regular and supplemental trend blocks were even more apparent in the proficiency metric for both mathematics and science. Since a longitudinal trend instrument cannot change in character and still provide meaningful comparisons across time, the supplemental trend blocks were not included in the assessment.

Table 13-8
Male-Female Differences in Mean Percentage Correct by Block, Mathematics Trend

Block	Age 9			Age 13			Age 17		
	Male	Female	Difference	Male	Female	Difference	Male	Female	Difference
Regular									
M1	63	63	0	66	65	1	65	63	2
M2	62	62	0	61	59	2	67	64	3
M3	68	69	-1	65	64	1	55	52	3
Total	—	—	-1	—	—	4	—	—	8
Supplemental									
MM	63	61	2	70	70	0	71	69	2
MN	54	57	-3	63	65	-2	61	60	1
MO	57	57	0	50	52	-2	59	59	0
Total	—	—	-1	—	—	-4	—	—	3

13.1.3 Estimation of Item Parameters

The first step in the scaling process was the estimation of item parameters for the trend items. Using the NAEP version (Rogers & Nelson, 1990) of the BILOG program (Mislevy & Bock, 1982), this item calibration was performed separately for each of the three age groups, using combined data from several assessment years. This assures that item parameters will be similar for adjacent assessments so that year-to-year trends will not be distorted by abrupt changes in calibration. The calibration was performed on a subsample of all the available subjects, resulting in approximately 500 examinees in each assessment year for each item. As with the previous assessment, calculator items were excluded from the analysis. Because calculators have changed greatly since the start of the long-term trend, it was judged that calculator questions are no longer comparable across time. These items were left in the assessment, since excluding them would have changed the testing context. Initially the responses were not weighted for this part of the analysis because the weights were unavailable early in the analysis process. After convergence was reached without weights, student weights were added to the analysis, and final convergence was obtained.

Item parameters were estimated for items in blocks M1 through M3 at ages 9 and 13 separately using 1986, 1988, and 1990 data. Parameters were estimated for items in blocks M1 through M3 given to the age 17 trend samples using 1986 and 1990 data. Items were examined for lack of fit with the data. Those that exhibited extreme violation of IRT assumptions (i.e., did not have monotonically increasing item characteristic curves) were deleted from the analysis. Other items were deleted because they were calculator items, which are not considered part of the regular assessment. These items appear in Table 13-9, 13-10, and 13-11. As a result of these deletions, 55 items were scaled for age 9, 80 items were scaled for age 13, and 71 items were scaled for age 17. Of the 153 noncalculator items that were part of the assessment, seven items (5 percent) were excluded due to poor fit with the data. A list of the items scaled for each of the ages, along with their item parameter estimates, appears in Tables E-15, E-16, and E-17 in Appendix E.

13.1.4 Derived Background Variables

In the trend analysis, all derived variables based upon background questions were used both for conditioning and in reporting (to define subgroups). Derived reporting and conditioning variables are described in Appendix B.

13.1.5 Generation of Plausible Values

The generation of plausible values was conducted independently for each age group. In this approach we used student background information to condition item responses in order to more accurately estimate student abilities. Rogers' (1991) enhancement of the MGROUP computer program (Sheehan, 1985) was used to combine NAEP-BILOG item parameters with weighted item responses and background variables to produce posterior ability estimates called plausible values. Because there were fewer background variables available for the trend samples, fewer conditioning variables were used in the creation of the plausible values on the

Table 13-9
Items Deleted from the Age 9 Mathematics Trend Analysis

Booklet	Block	Item	Reason for Exclusion
91	M1	N252601	Was deleted in prior assessment
		N262502	Bad fit with data in 1990
92	M3	N268221	Calculator item*
		N276021	Calculator item
		N276022	Calculator item
		N276821	Calculator item
		N276822	Calculator item
		N276823	Calculator item
		N277621	Calculator item
		N277622	Calculator item
		N277623	Calculator item
		N284021	Calculator item
		N284022	Calculator item

* Note that all calculator items are deleted from the analysis.

Table 13-10
Items Deleted from the Age 13 Mathematics Trend Analysis

Booklet	Block	Item	Reason for Exclusion
91	M1	N262502	Bad fit with data in 1990
93	M2	N261601	Bad fit with data in 1990
92	M3	N264521	Calculator item*
		N259921	Calculator item
		N276821	Calculator item
		N276822	Calculator item
		N276823	Calculator item
		N278921	Calculator item
		N278922	Calculator item
		N278923	Calculator item
		N278924	Calculator item
		N278925	Calculator item
		N280621	Calculator item
		N280622	Calculator item
		N280623	Calculator item
		N280624	Calculator item
		N280625	Calculator item
		N280626	Calculator item

* Note that all calculator items are deleted from the analysis.

Table 13-11
Items Deleted from the Age 17 Mathematics Trend Analysis

Booklet	Block	Item	Reason for Exclusion
84	M1	N282801	Bad fit with data in 1990
		N285701	Was deleted in prior assessment
84	M2	N266801	Was deleted in prior assessment
		N255301	Bad fit with data in 1990
85	M3	N259921	Calculator item*
		N264321	Calculator item
		N264521	Calculator item
		N257921	Calculator item
		N276821	Calculator item
		N276822	Calculator item
		N276823	Calculator item
		N278921	Calculator item
		N278922	Calculator item
		N278923	Calculator item
		N278924	Calculator item
		N278925	Calculator item
		N280621	Calculator item
		N280622	Calculator item
		N280623	Calculator item
		N280624	Calculator item
		N280625	Calculator item
		N280626	Calculator item
		N285321	Calculator item

* Note that all calculator items are deleted from the analysis.

trend scale than on the cross-sectional scale. There were 45 contrasts in the conditioning model at age 9, 48 at age 13, and 55 at age 17. Appendix F gives the codings for the conditioning variables (Table F-5) and the estimated conditioning effects (Tables F-21, F-22, and F-23) for the three age groups. The estimated conditioning effects in the tables are expressed on the scale of the original calibration (i.e., the theta scale). A check was made on the distributions of the plausible values for each age, including inspection of the whole group and subgroup means and standard deviations.

13.1.6 The Final Proficiency Scale

Since the cross-sectional scale is arbitrary, comparisons with previous assessments will be sensible only if the scale is linearly transformed to a meaningful metric. This is done by linking the 1990 scales to previous trend scales. The 1990 data needed to be transformed to compensate for linear changes in the scale due to having new item parameters and new NAEP-MGROUP conditioning parameters in 1990. This was accomplished by first reestimating the 1986 student abilities using 1990 item and NAEP-MGROUP parameters. The new 1986 ability estimates were then equated to the old 1986 ability estimates by matching the first two moments (i.e. the mean and standard deviation). The constants for this transformation were then applied to the 1990 data. The transformation equations that resulted are:

$$\text{Age } 9\text{: } \theta_{proficiency} = 34.37 \bullet \theta_{calibrated} + 226.18$$

$$\text{Age } 13\text{: } \theta_{proficiency} = 31.84 \bullet \theta_{calibrated} + 270.43$$

$$\text{Age } 17\text{: } \theta_{proficiency} = 30.25 \bullet \theta_{calibrated} + 303.55\ ,$$

where $\theta_{proficiency}$ denotes an individual's value on the final transformed scale of the 1990 data and $\theta_{calibrated}$ denotes an individual's value on the original 1990 theta scale. Overall summary statistics for the trend samples are given in Table 13-12.

Table 13-12
Means and Standard Deviations on the Mathematics Trend Proficiency Scale

Age	Assessment Year	All Five Plausible Values	
		Mean	S. D.
9	1986 1990	221.7 229.6	34.0 32.9
13	1986 1990	269.0 270.4	30.8 31.1
17	1986 1990	302.0 304.6	31.0 31.1

13.2 CROSS-SECTIONAL DATA ANALYSIS

The cross-sectional analysis included two assessments: the main focused-BIB assessment and the special mathematics assessment. The two assessments differ in measurement objectives, which resulted in different modes of assessment and different items.

The data from the main focused-BIB assessment of mathematics (from samples 9[Math-MainP, 13[Math-MainP], and 17[Math-MainP]) were used for cross-sectional analyses comparing the levels of mathematics achievement for various subgroups of the 1990 target populations. The main assessment included three student cohorts: students who were either in the fourth grade or 9 years old, students who were either in the eighth grade or 13 years old, and students who were either in the twelfth grade or 17 years old. The birth date ranges for age-eligible students were based on the 1980, 1976, and 1972 calendar years respectively for ages 9, 13 and 17. The sampled students in each of these three cohorts were assessed either in the winter or the spring. The samples in the main assessment are listed in Table 13-1.

The pool of items used in the 1990 mathematics assessment contained a range of open-ended and multiple-choice questions measuring performance on sets of objectives documented in *Mathematics Objectives: 1990 Assessment* (NAEP, 1988). The objectives framework is described in Chapter 2. A total of 282 distinct mathematics items addressing these objectives was administered in 1990 using the focused-BIB design to allocate the items to the assessed students. The items were classified into five categories based on their content: numbers and operations; measurement; geometry; data analysis, statistics, and probability; and algebra and functions. These five categories of items constituted the subscales used in 1990. Each booklet contained relatively few items from each of the five categories.

In the main samples, each student was administered a booklet containing three blocks of mathematics cognitive items, a block of background questions common to all booklets for a particular age/grade level, and a block of mathematics-related background questions common to all mathematics booklets for a particular age/grade level. Seven blocks of mathematics cognitive questions were administered at each age/grade level in a total of seven booklets for each level. (See Chapter 4 for more information about the blocks and booklets.) The seven blocks were not intended to be parallel measuring instruments. For example, two blocks contained only the items designed for calculator usage, and one block contained items for ruler and protractor usage. In addition, the number of items sampled from the five categories were not necessarily similar among the seven blocks. Both age-selected and grade-selected students contributed to the cross-sectional scaling. However, the "grade-only" portion of the main focused-BIB mathematics samples (whether the time of assessment was winter or spring) contributed to the means and percentages of the cross-sectional results that are reported in *The State of Mathematics Achievement: NAEP's 1990 Assessment of the Nation and the Trial Assessment of the States*. There were 275 unique items scaled for all three age/grade levels. Taking into account overlapping items, there were 77 multiple-choice and 28 open-ended items at grade 4, 101 multiple-choice and 35 open-ended items at grade 8, and 108 multiple-choice and 35 open-ended items at grade 12. The number of items that overlapped age/grade levels are listed in Table 13-13. Numbers of items in each subscale by block and by booklet are given in Tables E-3 and E-4 in Appendix E.

Table 13-13
Number of Scaled Mathematics Items Common Across Grades, by Subscale

Grade	Numbers and Operations	Measurement	Geometry	Data Analysis, Statistics and Probability	Algebra and Functions	Total
4 only	33	12	7	0	7	59
8 only	21	7	9	4	8	49
12 only	20	12	13	10	24	79
4 and 8 only	8**	4	5**	3*	4	24
4 and 12 only	0	1	0	0	0	1
8 and 12 only	6**	6	10	10**	10**	42
4, 8, and 12	11***	3	2	2*	3	21
Total	99	45	46	29	56	275

* These items were *not* calibrated at grade 4.
** One item in this scale was calibrated separately by grade.
*** Two items in this scale were calibrated separately by grade.

The nine items from the category of data analysis, statistics, and probability administered to the age 9/grade 4 group were distributed over seven blocks. As a result, the numbers of items per examinee were too few to provide reliable proficiency estimates of that scale for age 9/grade 4. Therefore, the subscale was deleted for the age 9/grade 4 analysis, and items in this scale that were given only to the grade 4/ age 9 group were dropped from the analysis. Table 13-13 contains items that were included in the analysis and excludes the few items that were administered but not included due to various reasons described in section 13.2.4.

13.2.1 Special Mathematics Assessment

The special samples 9[Math-MainT], 13[Math-MainT], and 17[Math-MainT] were taken from the same populations as the main mathematics assessment in order to focus on students' abilities in estimation and higher-order thinking skills.

Each age/grade sample of approximately 3,150 students was administered one booklet of four blocks appropriate for his or her age/grade level. Every booklet had one nonpaced block in common with the main assessment and a 45-minute audiotaped assessment consisting of three blocks of some combination of estimation and higher-order thinking skills items. Each booklet contained two sets of noncognitive questions; a block of background questions common to all main focused-BIB booklets for a particular age/grade level, and a block of mathematics-related background questions common to all main focused-BIB mathematics booklets for a particular age/grade level. The number of estimation and higher-order thinking skills items for each age/grade booklet is shown in Table 13-14.

Table 13-14
Special Mathematics Assessment Blocks and Items

Sample	Estimation Items		Higher-order Thinking Skills Items	
	Number	Blocks*	Number	Blocks
Age 9/Grade 4	20	MJ	14	MK, ML
Age 13/Grade 8	46	MJ, MK	8	ML
Age 17/Grade 12	46	MJ, MK	13	ML

* Age 13/Grade 8 and Age 17/Grade 12 received identical estimation blocks.

Although the estimation items were classified into the five subscales found in the main assessment, the number of items in a single subscale was too small to support stable linking to the main assessment separately by subscale. It was therefore decided to treat estimation items as if they were a single subscale. Consequently, the estimation proficiency distribution was linked to the single-scale composite distribution of the main assessment. The resulting estimation proficiency scale was transformed to have a mean of 250.5 and a standard deviation of 50. Since it was necessary to compare the national results with the results from the Trial State Assessment (where the estimation items were not presented) the estimation scale was not included in the overall composite scale, which was based only on the five main mathematics subscales.

As the item analyses were being carried out, it was observed that not only were items in the estimation scale and higher-order thinking skills scale very different from each other, but items in both scales were different from items in the main assessment in terms of their response distributions. For both types of items there were higher numbers of omits than for the regular items, especially for Black and Hispanic students. The omit rate was particularly high for higher-order thinking skills items. For example, the rate for higher-order thinking skills items ranged from 1 to 20 percent for 17-year-old White students, but from 4 to 29 percent for 17-year-old Black students. An even greater problem was the possible floor effect on ability estimates caused by high numbers of minorities who answered none of the higher-order thinking skills items correctly. As shown in Table 13-15, at age 17, 15 percent of the White students answered all higher-order thinking skills items incorrectly, while 33 percent of Hispanic students and 42 percent of Black students answered them all incorrectly. With so few minorities getting any of the items correct there is a floor effect such that there is no information for distinguishing among the higher-order thinking skills abilities for many of the students. Because of the high omit rates and low number of students getting any correct responses among minorities, it was decided that the higher-order thinking skills items would *not* be combined with estimation items, and the IRT-based analysis would not be applied to these items. Instead they were subjected to a classical analysis consisting of estimates such as average proportion correct and item/total-test statistics.

The number of overlapping items for the age/grade levels are listed in Table 13-16. Numbers of items in the estimation subscale by block and by booklet are given in Table E-5 in Appendix E. The tabled values reflect only those items included in the final analysis.

Table 13-15

Percentage with All Items Incorrect for Higher-order Thinking Skills Blocks by Grade and Race/Ethnicity

Grade	Race/Ethnicity	Percentage with All Items Incorrect	
		Block MK	Block ML
4	White	1	16
	Hispanic	2	41
	Black	3	50
8	White	—	2
	Hispanic	—	8
	Black	—	12
12	White	—	15
	Hispanic	—	33
	Black	—	42

Table 13-16

Number of Estimation and Higher-order Thinking Skills Items Common Across Grades

Grade	Number of Items	
	Estimation	Higher-order Thinking Skills
4 only	10	7
8 only	0	0
12 only	0	11
4 and 8 only	0	6
4 and 12 only	0	0
8 and 12 only	36	1
4, 8, and 12	10	1
Total	56	26

13.2.2 Item Analysis

The next sections contain in some detail a description of the analysis performed using the main focused-BIB sample data. As with the trend analysis, the process began with an examination of items within blocks. Open-ended items were dichotomously scored, items were grouped by content domains, and derived background variables were calculated. The estimation of item parameters for the five mathematics subscales was completed, followed by the generation of multivariate plausible values. Finally, the plausible values were transformed to the final proficiency scale.

Tables 13-17, 13-18, and 13-19 show the number of items, mean proportion correct, mean r-biserial, and alpha reliability for each block administered at each age/grade level for the main assessment.[2] These values were calculated for the dichotomously scored multiple-choice and open-ended items within a block if they were used in the scaling process. The table also gives the number of students who were administered the block and the percent not reaching the last item in the block. These numbers only include the students in the grade-only portion of the samples that contributed to the summary statistics provided in *The State of Mathematics Achievement*. Student weights were used, except for the sample sizes. The results for the blocks administered to each age/grade level indicated that the blocks differ in number of items, average difficulty, reliability, and percent not reaching the last item. Preliminary item analyses for all items within a block were performed before scaling; the results shown here indicate the characteristics of the items that contributed to the final scale.

Table 13-20 shows estimates for special sample items that parallel the statistics described above for the main assessment. These blocks are presented separately because they were not analyzed together with other subscales. Moreover, the results from the special sample items were never part of the composite scale results.

Tables 13-17, 13-18, and 13-19 also contain information about the effect of the position of blocks within booklets on the average percent correct for items within each block presented to the focused-BIB samples for each grade. The averages for the grade-only portion of the focused-BIB samples show that the order of blocks within booklets did not have a large or consistent effect on proficiency in the mathematics focused-BIB assessment. The items with their classifications are listed in the data appendix of *The State of Mathematics Achievement*. The items are listed by subscale in Tables E-3 and E-4 of Appendix E.

13.2.3 Scoring the Open-ended Items

As indicated earlier, the mathematics assessment included many open-ended items. These items were included in the scaling process, but were also analyzed separately. Chapter 7 contains the means and ranges for percent agreement between raters for the items as they were

[2] It should be noted that each block contained items from five subscale categories and all analyses for the report were carried out at first by each subscale separately and later combined to form the composite scale. Hence, the block level statistics presented here do not necessarily correspond directly to the composite results.

Table 13-17

Descriptive Statistics for Item Blocks by Position Within Booklet and Over All Occurrences
Mathematics Cross-sectional Sample, Grade 4

Statistic	Block Position	Block MC	MD	ME	MF	MG	MH	MI
Number of scaled items	——	19	14	11	17	15	14	15
Number of scaled open-ended items	——	6	0	11	4	0	1	6
Unweighted sample size	1	939	913	923	904	913	928	937
	2	930	934	907	918	897	906	923
	3	904	921	928	934	904	917	897
	ALL	2773	2768	2758	2756	2714	2751	2757
Average weighted proportion correct	1	.63	.43	.38	.40	.60	.59	.49
	2	.64	.42	.35	.41	.58	.58	.50
	3	.62	.43	.34	.38	.59	.58	.48
	ALL	.63	.43	.36	.40	.59	.58	.49
Average weighted r-biserial	1	.59	.51	.68	.54	.57	.55	.57
	2	.58	.53	.70	.54	.59	.53	.58
	3	.58	.52	.69	.53	.60	.55	.57
	ALL	.58	.52	.69	.54	.59	.54	.57
Weighted alpha reliability	1	.77	.58	.72	.70	.70	.66	.71
	2	.75	.61	.73	.69	.73	.63	.71
	3	.75	.58	.71	.67	.74	.65	.67
	ALL	.76	.59	.72	.69	.72	.65	.70
Weighted proportion of students attempting last item	1	.95	.96	.76	.70	.89	.79	.51
	2	.96	.97	.76	.76	.91	.88	.65
	3	.96	.93	.77	.88	.94	.87	.59
	ALL	.96	.95	.76	.78	.91	.84	.58

Table 13-18

Descriptive Statistics for Item Blocks by Position Within Booklet and Over All Occurrences
Mathematics Cross-sectional Sample, Grade 8

Statistic	Block Position	Block MC	MD	ME	MF	MG	MH	MI
Number of scaled items	——	23	21	16	21	18	18	19
Number of scaled open-ended items	——	4	0	16	5	1	2	7
Unweighted sample size	1	916	932	917	925	947	928	903
	2	903	916	929	916	924	947	926
	3	938	920	897	910	927	911	917
	ALL	2757	2768	2743	2751	2798	2786	2746
Average weighted proportion correct	1	.67	.56	.52	.65	.43	.49	.52
	2	.66	.55	.51	.65	.43	.50	.52
	3	.68	.54	.50	.61	.41	.47	.52
	ALL	.67	.55	.51	.64	.42	.49	.52
Average weighted r-biserial	1	.58	.53	.67	.64	.58	.56	.61
	2	.58	.56	.69	.66	.58	.59	.58
	3	.59	.51	.69	.68	.59	.58	.62
	ALL	.58	.53	.68	.66	.58	.58	.60
Weighted alpha reliability	1	.81	.73	.81	.84	.75	.73	.80
	2	.81	.78	.83	.85	.74	.75	.77
	3	.81	.72	.83	.86	.76	.74	.81
	ALL	.81	.75	.83	.85	.75	.74	.79
Weighted proportion of students attempting last item	1	.99	.91	.89	.86	.95	.68	.59
	2	.98	.93	.91	.85	.96	.74	.68
	3	.98	.95	.93	.89	.97	.78	.65
	ALL	.98	.93	.91	.87	.96	.73	.64

Table 13-19

Descriptive Statistics for Item Blocks by Position Within Booklet and Over All Occurrences
Mathematics Cross-sectional Sample, Grade 12

Statistic	Block Position	Block						
		MC	MD	ME	MF	MG	MH	MI
Number of scaled items	—	23	22	17	19	21	21	20
Number of scaled open-ended items	—	5	0	17	3	3	4	3
Unweighted sample size	1	904	915	893	902	885	913	893
	2	890	901	909	891	900	882	910
	3	881	908	890	895	906	888	896
	ALL	2675	2724	2692	2688	2691	2683	2699
Average weighted proportion correct	1	.66	.66	.51	.56	.52	.51	.40
	2	.67	.67	.53	.56	.52	.51	.41
	3	.66	.65	.53	.55	.53	.50	.39
	ALL	.66	.66	.52	.55	.52	.51	.40
Average weighted r-biserial	1	.67	.58	.72	.58	.66	.63	.58
	2	.68	.59	.72	.60	.63	.62	.59
	3	.66	.58	.74	.59	.66	.63	.57
	ALL	.67	.59	.73	.59	.65	.63	.58
Weighted alpha reliability	1	.85	.77	.84	.75	.85	.80	.75
	2	.86	.77	.83	.78	.83	.80	.76
	3	.85	.78	.85	.75	.85	.80	.71
	ALL	.85	.77	.84	.76	.84	.80	.74
Weighted proportion of students attempting last item	1	.99	.88	.73	.69	.72	.50	.52
	2	.98	.89	.74	.67	.70	.59	.64
	3	.99	.88	.72	.78	.69	.56	.60
	ALL	.99	.88	.73	.71	.71	.55	.59

Table 13-20

Descriptive Statistics for Special Item Blocks
Mathematics Special Sample

Statistic	Block		
	MJ (Estimation)	MK (Higher-order Thinking Skills)	ML (Higher-order Thinking Skills)
Grade 4			
Number of items in a block	20	7	7
Number of scaled items	19	0	0
Number of open-ended items	0	7	6
Unweighted sample size	2435	2435	2435
Average weighted proportion correct	.56	.58	.25
Average weighted r-biserial	.42	.73	.80
Weighted alpha reliability	.51	.58	.64
Grade 8			
Number of items in a block	22	24	8
Number of scaled items	21	24	0
Number of open-ended items	0	0	7
Unweighted sample size	2415	2415	2415
Average weighted proportion correct	.57	.61	.50
Average weighted r-biserial	.49	.52	.76
Weighted alpha reliability	.68	.76	.71
Grade 12			
Number of items in a block	22	24	13
Number of scaled items	21	24	0
Number of open-ended items	0	0	13
Unweighted sample size	2371	2371	2371
Average weighted proportion correct	.66	.75	.30
Average weighted r-biserial	.55	.64	.81
Weighted alpha reliability	.75	.82	.79

originally scored. The right/wrong scoring of the categories of responses for the items and the number of responses that were rescored for each item are indicated in Tables G-2, G-3, and G-4 in Appendix G. The percent agreement for the raters and Cohen's Kappa, a reliability estimate appropriate for items that are dichotomized, are also given in the tables. The sample sizes listed in the tables correspond to the samples used in calculating the rater reliability.

In general, the rater reliability of the scoring for dichotomized responses was quite high. Cohen's Kappa reliabilities ranged over items from .79 to .99 for age 9, from .87 to .99 for age 13, and from .79 to .99 for age 17.

13.2.4 Estimation of Item Parameters

The NAEP-BILOG computer program was used to estimate the item parameters of the three-parameter IRT model for the entire 275 items, using a random subsample of 12,905 of the 25,572 students in the main assessment samples. Items presented to the three age/grade groups were calibrated together using NAEP-BILOG in a mode that estimates parameters for the five subscales separately, while allowing for different means and variances across grades. The number of subjects responding to an item ranged from 963 to 5,740, with all but one item having greater than 1,000 responses. In general, identical item parameters were used for the different age/grade levels and for winter and spring samples. The appropriateness of the use of the identical parameters across age/grade and winter/spring was examined by comparing the fit of the empirical item response functions against the estimated IRT item response functions. This was done for the six groups of students from all three age/grade levels and both winter and spring administrations. If IRT parameters did not fit the data, parameters specific to the age/grade and winter/spring subgroup were used. In the next step, sampling weights were applied and a final item estimation was run to convergence. (See Chapter 11 for further descriptions of the scaling process.)

Several items were deleted from the IRT analysis or were allowed to have different item parameters for separate age/grade groups due to lack of fit to the model. Of the 275 total items, 17 (6 percent) received special treatment. These items are listed in Table 13-21, along with the reason for special treatment. If items had item response functions (IRFs) that were nonmonotonic, the items were deleted from scaling. If item response functions for items presented to the various age/grades or administrations (winter or spring) differed for some of the groups, the item was treated as different items for the groups. As it turned out, item parameters never differed from winter to spring administrations.

The IRT parameters for estimation items were also estimated using the NAEP-BILOG computer program. All items were estimated together on a calibration sample which consisted of a random sample of three age/grade groups. The calibration sample size was selected in order to ensure that the minimum number of responses for each item would be at least 1,000. The sample weights were used so that estimates would accurately reflect the composition of the actual population. Six subgroups, defined by three age/grade and two winter/spring subsamples, were given different prior ability distributions during item parameter estimation. After evaluating the estimated item response functions for the six subgroups, it was decided that three items would have varying item parameters for different subgroups.

Table 13-21

Items from the Mathematics Cross-sectional Analysis Receiving Special Treatment

Item	Grade	Block	Content Area	Treatment	Reason for Treatment
N255701	8, 12	MC	Algebra and Functions	Separate estimates for grades 8 and 12	Differing IRFs
M020001	4, 8	ME	Numbers and Operations	Separate estimates for grades 4 and 8	Differing IRFs
M028831	8	ME	Measurement	Dropped from grade 8	Nonmonotonic
M019801	4, 8	MF	Geometry	Separate estimates for grades 4 and 8	Differing IRFs
M022001	4, 8, 12	MF	Numbers and Operations	Separate estimates for grade 4	Differing IRFs
M024901	12	MF	Measurement	Dropped from grade 12	Nonmonotonic
M025301	12	MF	Numbers and Operations	Dropped from grade 12	Nonmonotonic
M013901	4	MG	Data Analysis, Statistics and Probability	Dropped from grade 4	Subscale 4 was dropped for grade 4
M014101	4	MG	Data Analysis, Statistics and Probability	Dropped from grade 4	Subscale 4 was dropped for grade 4
M015001	4	MG	Data Analysis, Statistics and Probability	Dropped from grade 4	Subscale 4 was dropped for grade 4
M012431	8, 12	MH	Numbers and Operations	Separate estimates for grades 8 and 12	Differing IRFs
M012631	8, 12	MH	Data Analysis, Statistics and Probability	Separate estimates for grades 8 and 12	Differing IRFs
M202831	4, 8, 12	MH	Numbers and Operations	Separate estimates for grade 12	Differing IRFs
M250231	4	MH	Data Analysis, Statistics and Probability	Dropped from grade 4	Subscale 4 was dropped for grade 4
M032001	4, 8, 12	MJ	Estimation	Separate estimates for all 3 grades	Differing IRFs
M032101	4, 8, 12	MJ	Estimation	Dropped from grades 8 and 12	Nonmonotonic
M032801	4, 8, 12	MJ	Estimation	Dropped from grade 4	Nonmonotonic

BEST COPY AVAILABLE

Tables E-18 through E-23 in Appendix E list the estimated item parameters that were produced by the NAEP-BILOG program. Because of the indeterminacy of the IRT scale, the origin and size of the scale were set provisionally by standardizing the combined distribution of the calibration samples of six groups of examinees (three age groups by winter and spring samples) to have a mean of zero and a variance of one.

13.2.5 Derived Background Variables

Derived variables based upon background questions were used for two purposes: as conditioning variables and as reporting variables used to define subgroups. Some of these variables are common to all the subject areas; others are specific to the 1990 mathematics assessment. Derived variables used for conditioning and reporting are described in Appendix B.

13.2.6 Generation of Plausible Values

For the entire sample, multivariate plausible values for subscales were generated for each age/grade group separately using the NAEP-MGROUP program (Rogers, 1991; Sheehan, 1985). Final student weights were used at this stage of the analysis. Instead of using selected background variables for conditioning variables (as had been done in the past), principal components of the background variables were used. The principal components used accounted for 90 percent of the variance of the original conditioning variables. Principal components were used to remedy problems of extreme collinearity among some of the original conditioning variables. For the estimation scale, univariate plausible values were generated for each age/grade group separately; otherwise, the same procedures were followed as for the main administration.

The codings of the original mathematics-specific conditioning variables, before principal components were calculated, are presented in Appendix F in Table F-4. (For age 17/grade 12, the "modal age, > modal grade" category was deleted from the age-by-grade variable, because students above grade 12 were not sampled.) NAEP-MGROUP creates predictive distributions of proficiencies by combining information from item responses of individuals and information from linear regression of proficiency on conditioning variables.

The proportion of variance of each original conditioning variable accounted for by the principal components included in the conditioning model is listed in Tables F-15, F-17, and F-19. The estimated conditioning effects for the principal components of the three samples defined by the three age/grade groups are given in Appendix F in Tables F-16, F-18, and F-20. The values of the conditioning effects are expressed in the metrics of the original calibration scale. Definitions of derived conditioning variables are given in Appendix B.

13.2.7 The Transformation of the Proficiency Scale for Reporting and the Formation of the Composite Scale

Like all IRT scales, the mathematics subscales have a linear indeterminacy that may be resolved by an arbitrary choice of the origin and unit-size in each given subscale. Following the

convention established in the previous NAEP assessments, the subscale mean and the standard deviation were set to 250.5 and 50.0. The linear indeterminacies among the subscales was resolved by transforming the subscale means and variances of three age/grade samples combined together to the 250.5, 50.0 metric. As a result, all of the subscales that spanned all three age/grade samples are now on the same common scale. One subscale was dropped due to inadequate number of items for age 9/grade 4. For that scale, the means and variances of the two remaining age/grade samples were matched to the weighted means of other subscales were transformed to the desired metric.

For each subscale the same linear transformation was applied to the proficiencies of all age/grade samples. Table 13-22 shows the coefficients of the linear transformations used to transform the subscales from their original units (calibrating scale) to the final composite proficiency scale.

Table 13-22
Coefficients of the Linear Transformations of the Subscales
from Calibrating Scale to the Units of the Reporting Proficiency

Subscale	Intercept	Slope
Numbers and Operations	251.72	50.35
Measurement	252.59	49.99
Geometry	252.58	49.15
Data Analysis, Statistics, and Probability	274.08	44.47
Algebra and Functions	252.01	50.34

A linear indeterminacy of the estimation scale was resolved by setting the mean and the standard deviations to be 250.5 and 50.0 for the proficiencies of the three age/grade samples combined. The transformation constants used for the above purpose are 249.523 for the intercept and 49.617 for the slope.

While multiple proficiency scales provide useful and very revealing information about the relative relationships among subpopulations, a single index to summarize overall performance is useful and communicative. For that reason, a mathematics composite was defined as a weighted average of the results across subscales. Not all subscales apply to all age/grade samples nor does the importance associated with each subscale remain the same across all age/grade samples. Therefore, the weights assigned to compute the average of the estimated subscale proficiencies differ by age/grade. The assigned weights reflect the relative importance of subscales for a particular age/grade as specified in *Mathematics Objectives, 1990 Assessment* (NAEP, 1988). This is a nearly optimal weighting of the subscales in terms of the precision of the resulting composite. This is because the actual weighting corresponded closely to the scale information, and scale information weighting produces the most precise composite. The definition of weights for the composite in each age/grade is given in Table 13-23.

Table 13-23
Defining Weights for the Mathematics Composite by Age/Grade

Subscale	Age 9/Grade 4	Age 13/Grade 8	Age 17/Grade 12
Numbers and Operations	50	30	25
Measurement	22	15	15
Geometry	17	20	20
Data Analysis, Statistics, and Probability	0	15	15
Algebra and Functions	11	20	25

Finally, it is necessary to caution that, although the mathematics composite is expressed in seemingly the same units as the 1986 mathematics scale, it is not appropriate to compare scores on the mathematics composite with scores from the other cross-sectional assessments. The transformation chosen to resolve the linear indeterminacies in the mathematics composite is a convenient transformation, but is only one of a conceptually infinite number of such transformations that could have been chosen, any one of which would have provided equivalent information about the relative standings of subgroups of the population in terms of their abilities in mathematics. Because there was no link, real or implied, in the construction of the mathematics composite and the subscales to either science or reading assessments or to the previous mathematics assessments, the comparison of students' mathematics proficiencies to students' proficiencies on other subjects is devoid of meaning.

Summary statistics for the composite samples are given in Table 13-24.

Table 13-24
Means and Standard Deviations on the Mathematics Cross-sectional Composite Proficiency Scale

Grade	All Five Plausible Values	
	Mean	S. D.
4	215.8	28.1
8	265.0	32.4
12	295.3	33.3

13.2.8 Partitioning of the Estimation Error Variance

For each scale within each grade, the error variance of the transformed proficiency means was partitioned according to the procedure described in Chapter 11. The variance is partitioned into two parts; the proportion of error variance due to sampling students (sampling variance) and the proportion of error variance due to the fact that proficiency, Θ, is a latent variable that is estimated rather than observed. Table 13-25 contains estimates for the sampling variance U^* and estimates of variance due to the latency of Θ which equals $(1+M^{-1})B$, where $M=5$, the number of imputations, and B is the variance among the means of the five plausible

values. The table also contains estimates of the total error variance, V, the proportion of error variance due to sampling students and finally, the proportion of error variance due to the latent nature of Θ.

Table 13-25
Estimation Error Variance and Related Coefficients for the Mathematics Cross-sectional Assessment

Grade	Scale	U^*	$(1+5^{-1})B$	V	Proportion of Variance Due to... Student Sampling: U^*/V	Latency of θ: $(1+5^{-1})B/V$
4	Composite	0.45	0.03	0.47	0.95	0.05
	Numbers & operations	0.56	0.09	0.65	0.86	0.14
	Measurement	0.51	0.20	0.71	0.71	0.29
	Geometry	0.55	0.03	0.58	0.95	0.05
	Algebra & functions	0.36	0.09	0.45	0.81	0.19
8	Composite	1.08	0.01	1.09	0.99	0.01
	Numbers & operations	1.07	0.03	1.10	0.97	0.03
	Measurement	1.46	0.02	1.48	0.98	0.02
	Geometry	1.06	0.02	1.08	0.99	0.01
	Data analysis, stat., & prob.	1.58	0.10	1.68	0.94	0.06
	Algebra & functions	1.01	0.05	1.06	0.95	0.05
12	Composite	1.09	0.01	1.11	0.99	0.01
	Numbers & operations	0.88	0.09	0.97	0.91	0.09
	Measurement	1.19	0.01	1.21	0.99	0.01
	Geometry	1.59	0.06	1.65	0.96	0.04
	Data analysis, stat., & prob.	1.13	0.06	1.19	0.95	0.05
	Algebra & functions	1.15	0.04	1.19	0.97	0.03

13.2.9 Mathematics Teacher Questionnaire

Teachers of fourth- and eighth-grade students assessed in mathematics were surveyed. Variables derived from the questionnaire were used in the conditioning models for the age 9/grade 4 and the age 13/grade 8 samples, along with a variable that indicated whether a student record had been matched with a teacher record, so that means for subgroups defined by these variables could be compared with no bias. Questionnaires were received from 393 fourth-grade and 597 eighth-grade teachers. Of the 6,467 fourth-grade students in the main focused-BIB sample, 5,436 (84 percent) were matched with both parts of the teacher questionnaire and 404 (6.2 percent) were matched with only the first part of the questionnaire. Of the 6,473 eighth-grade students in the main focused-BIB sample, 4,908 (76 percent) were matched with both parts of the teacher questionnaire and 714 (11 percent) were matched with only the first part of the questionnaire. Thus, 90 percent of the fourth graders and 87 percent of the eighth graders were matched with at least the background information about their mathematics teachers.

13.2.10 Mantel-Haenszel DIF Analyses

After item parameters had been estimated, a differential item functioning (DIF) analysis of the main-assessment mathematics items was done in order to aid test developers in future mathematics assessments. Sample sizes were large enough to compare male and female students, White and Black students, and White and Hispanic students using the Mantel-Haenszel procedure described in Chapter 9 to identify items that should be examined more closely for possible bias ("C" items). Table H-2 in Appendix H identifies those items that were categorized as "C" items in at least one analysis. Of the 275 scaled items, 23 (8 percent) showed significant DIF. Of these, 15 items showed DIF between racial/ethnic groups and 11 items showed DIF between males and females. The table also shows the block and subscale containing the item, and the grade and analysis for which the item was identified. Table 13-26 summarizes information about the identified items for each block.

Table 13-26
Numbers of "C" Items Favoring Each Group by Mathematics Block

Block	Grade	Number of Scaled Items in Block	Analysis: Male/Female: Male	Female	White/Black: White	Black	White/Hispanic: White	Hispanic
MC	4	19	0	0	0	0	0	0
	8	23	0	0	0	0	0	0
	12	23	1	1	0	0	0	0
MD	4	14	0	0	0	0	0	0
	8	21	0	0	0	0	0	0
	12	22	0	0	1	0	0	0
ME	4	11	0	1	0	0	0	0
	8	16	0	0	1	0	0	0
	12	17	2	0	2	0	1	0
MF	4	17	0	0	2	0	0	0
	8	21	0	0	0	0	0	0
	12	20	0	0	0	0	0	0
MG	4	18	0	0	0	0	0	0
	8	18	0	0	0	0	0	0
	12	21	1	0	1	0	0	0
MH	4	15	0	0	0	0	0	0
	8	18	1	0	0	1	0	0
	12	21	0	0	0	0	0	0
MI	4	15	0	0	0	0	0	0
	8	20	1	0	0	0	0	0
	12	20	0	0	1	0	0	0

13.2.11 Analysis of Dimensionality

As mentioned earlier, the cross-sectional assessment is multivariate with five subscales. To check the assumption of multidimensionality, a confirmatory factor analysis was performed. The result of the analysis is a set of factor correlations that are estimated assuming five factors (four for grade 4) corresponding to the five mathematics subscales. If all factor correlations were very high, this would indicate that there is little information in the five subscales that could not be captured by assuming a single scale, as was assumed with the trend analysis. Note that grade 4 had only four subscales; the following description, if amended for four subscales, would be true for grade 4.

The confirmatory factor analysis was performed on testlet scores for each booklet within each grade. A testlet is a small group of items of similar content, in this case a random half of the subscale items for a booklet. The seven booklets at each grade support seven replications of the analysis. Since there were five subscales per booklet and each subscale of items was randomly divided into two testlets of items, there were ten testlets for each booklet. Within each booklet, the number correct score was calculated for each testlet and polychoric correlations of testlet scores were used as the input for the analysis. A factor analytic model was posited such that each subscale was measured by two testlet scores. For each grade the correlations between the five subscale factors (four factors for grade 4) averaged over the seven booklets, are shown in Tables 13-27, 13-28, and 13-29. The factor correlations, which are estimated latent scale correlations, are quite large. This suggests that there is quite a bit of redundancy in the scales.

Although a general factor accounts for the vast majority of the variability in the proficiencies, a reanalysis of the 1990 cross-sectional mathematics data by Muthén (1991) indicates that some interesting variation among subgroups can be observed by employing specific scale factors that are independent of a general factor. Muthén showed that different subgroups (e.g., males and females) had different profiles of the percent of variability attributable to subscale specific factors. These percentages were quite small, however, averaging, for example, only 9 percent for both males and females.

13.2.12 Anchoring the Points on the Mathematics Proficiency Scale

Scale anchoring was devised to associate descriptive statements of a student's ability with a level on a continuum of proficiency. This was done successfully with the 1986 reading, mathematics, and science scales. The same technique was applied to the 1990 mathematics composite scale. Four levels—200, 250, 300, and 350—were selected on the scale and chosen as anchor points. Each level was defined by a description of the types of questions that most students attaining that level of proficiency would be able to answer correctly while most students at least one level lower would answer incorrectly. In this way each level was exemplified by typical benchmark items that describe a subset of abilities indicative of that level of proficiency.

The anchoring procedures employed weighted composite proficiency scores, which were derived from multivariate subscale proficiencies. The anchoring was based on the empirical proportion correct scores at each anchor point. The empirical proportion correct was calculated

Table 13-27
Average Correlations Between Factors, Grade 4

	Numbers and Operations	Measurement	Geometry	Algebra and Functions
Numbers and Operations	1.00			
Measurement	.95	1.00		
Geometry	.91	.93	1.00	
Algebra and Functions	.96	.95	.96	1.00

Table 13-28
Average Correlations Between Factors, Grade 8

	Numbers and Operations	Measurement	Geometry	Data Analysis, Statistics and Probability	Algebra and Functions
Numbers and Operations	1.00				
Measurement	.95	1.00			
Geometry	.86	.90	1.00		
Data Analysis, Statistics and Probability	.95	.95	.91	1.00	
Algebra and Functions	.92	.89	.88	.93	1.00

Table 13-29
Average Correlations Between Factors, Grade 12

	Numbers and Operations	Measurement	Geometry	Data Analysis, Statistics and Probability	Algebra and Functions
Numbers and Operations	1.00				
Measurement	.96	1.00			
Geometry	.96	.96	1.00		
Data Analysis, Statistics and Probability	.97	.90	.90	1.00	
Algebra and Functions	.93	.90	.93	.86	1.00

by selecting subjects in an *a priori* range of the composite proficiency score. For the selected subjects, the item responses were averaged. The ranges were set at 12.5 units from anchor points. For example, students who scored between 287.5 and 312.5 on the proficiency scale (within 12.5 units of 300), and also were administered a particular item, were used to estimate the conditional probability of the correct response on that item by calculating average proportion correct for an item. To avoid problems of instability of the estimated probabilities for very small numbers of respondents to an item, the average proportion was not defined if fewer than 100 students at a given proficiency range responded to the item. For further details on scale anchoring, see Beaton and Allen (in press).

In the scale anchoring process for the mathematics composite, NAEP identified sets of items from the 1990 assessment that were good discriminators between subjects at different proficiency levels (anchor point ranges). The guideline used to select such items was that students at any given level would have a proportion correct of at least .65 (but often higher) on these mathematics questions, while the students at the next lower level would have a much lower proportion correct (less than .50) using the criterion that the difference in proportions exceeds .30 between adjacent levels. In the case of the first anchor point, 200, the only criteria for selecting items was that a subjects at that proficiency level have a minimum average proportion correct of .65. Mathematics educators examined these sets of empirically selected items and used their expert judgment to characterize each proficiency level (anchor point), contrasting tasks at that level with those at the levels just above and below. As a check on the generalizability of the interpretation process, the proficiency levels were independently described by two groups of mathematics educators, each group consisting of 10 members. Upon comparing the results, both groups agreed that the two sets of interpretations were very similar and that either set would have appropriately described the anchor item information. The two groups then worked together to obtain the final interpretations.

The anchor levels were:

200 - Simple additive reasoning and problem solving with whole numbers;
250 - Simple multiplicative reasoning and two-step problem solving;
300 - Reasoning and problem solving involving fractions, decimals, percents, elementary geometric properties, and simple algebraic manipulations; and
350 - Reasoning and problem solving involving geometric relationships, algebraic equations, and beginning statistics and probability.

Chapter 14

DATA ANALYSIS FOR THE SCIENCE ASSESSMENT[1]

Nancy L. Allen

Educational Testing Service

This chapter describes the analyses performed on the responses to the cognitive and background items in the 1990 assessment of science. These analyses led to the results presented in *Trends in Academic Progress: Achievement of U.S. Students in Science, 1969-70 to 1990; Mathematics, 1973 to 1990; Reading, 1971 to 1990; and Writing, 1984 to 1990* (Mullis, Dossey, Foertsch, Jones, & Gentile, 1991) and in *The 1990 Science Report Card: NAEP's Assessment of Fourth, Eighth, and Twelfth Graders* (Jones, Mullis, Raizen, Weiss, & Weston, 1992). The emphasis of this chapter is on the methods and results of procedures used to develop the IRT-based scale scores that formed the basis of these reports. However, some attention is given to the analysis of open-ended items as reported in *The 1990 Science Report Card.* The theoretic underpinnings of the IRT and plausible value methodology described in this chapter are given in Chapter 11.

The objectives of the science analyses were to

- prepare scale values and perform all analyses necessary to produce a long-term trend report in science. The science trend line includes the years 1970, 1973, 1977, 1982, 1986, and 1990.

- prepare scale values for the cross-sectional analysis of the main focused-BIB science samples. The scaling of science entailed development of several subscales and an overall composite.

The student samples that were administered science items in the 1990 assessment are shown in Table 14-1. (See Chapters 1 and 3 for descriptions of the target populations and the sample design used for the assessment.) Data from the first three samples (Sci-MainP) were used in the cross-sectional analysis; data from the rest of the samples were collected for trend purposes.

Because they consisted of different items and objectives, the cross-sectional data from the 1990 main focused-BIB samples were scaled separately from the data from the bridge samples that contributed to the trends in science achievement. Accordingly, the trend and cross-

[1]Data analysis and scaling were performed by Steven Isham. Robert Mislevy and Kentaro Yamamoto consulted on IRT scaling and generation of plausible values.

sectional analyses are presented in separate sections. Section 14.1 pertains to the scaling of the data from the trend bridges; section 14.2 contains information about the scaling of the data from the main focused-BIB samples.

Table 14-1
NAEP 1990 Science Student Samples

Sample	Booklets	Mode	Cohort Assessed	Time of Testing	Age Definition	Modal Grade	Number Assessed
9 [Sci-MainP]	18-24	Print	Age 9/grade 4	Winter, spring	CY	4	8,418
13 [Sci-MainP]	15-21	Print	Age 13/grade 8	Winter, spring	CY	8	8,709
17 [Sci-MainP]	15-21	Print	Age 17/grade 12	Winter, spring	CY	12	8,445
9 [RMS-Br86]	91-93	Mixed	Age 9	Winter	CY	4	6,235
13 [RMS-Br86]	91-93	Mixed	Age 13	Fall	CY	8	6,649
17 [RMS-Br86]	61-66*	Print	Age 17/grade 11	Spring	Not CY	11	8,338
17 [MS-Br86]	84-85	Tape	Age 17	Spring	Not CY	11	4,411
9 [MS-BrLT]	94-95	Tape	Age 9	Winter	CY	4	4,134
13 [MS-BrLT]	94-95	Tape	Age 13	Fall	CY	8	4,455
17 [MS-BrLT]	94-95	Tape	Age 17	Spring	Not CY	11	4,402

* Only booklets 63-65 contain science items; 4,184 age 17/grade 11 students received these booklets.

LEGEND:

Sci	Science
RMS	Reading, mathematics, and science
MS	Mathematics and science
MainP	Main assessment, print administration
Br86	Bridge to 1986
BrLT	Bridge for long-term trend
Print	Printed administration
Tape	Audiotape administration
Mixed	Mathematics and science administered by audiotape, reading administered by print
CY	Calendar year: birthdates in 1980, 1976, and 1972 for ages 9, 13, and 17
Not CY	Age 17 only: birthdates between Oct. 1, 1972 and Sept. 30, 1973

14.1 TREND DATA ANALYSIS

The science trend results reported in *Trends in Academic Progress* are based on paced-tape administrations and occur at all of the age levels. The samples involved in the analysis were samples 9[RMS-Br86], 13[RMS-Br86], and 17[MS-Br86] in Table 14-1. For ages 9 and 13, the bridge booklets for these samples contained blocks of reading, mathematics, and science items. The science and mathematics blocks were paced by tape-recordings and the reading blocks were presented in print form. The age 17 bridge booklets contained only mathematics and science blocks, both presented by paced-tape recordings. All students received a block of common background questions, distinct for each age. Subject-area background questions were presented in the cognitive blocks. The booklets for the age 9 and age 13 samples (booklets 91-

93) are the same as those used for bridge assessments in 1986 and 1988. The booklets for the age 17 sample (booklets 84-85) are the same as those used for the 1986 bridge assessment. The booklets and the blocks within those booklets are listed in Tables 4-7, 4-8, and 4-9 of Chapter 4. Additional information about all of the items in these blocks is in Tables 4-10, 4-11, and 4-12 of that chapter. This chapter includes specific information about the trend items that were scaled.

Table 14-2 clarifies the relationships between the 1990 trend samples and samples from previous years. For ages 9 and 13, the paced-tape bridge-to-1986 samples allow direct comparisons with 1988 samples (a year in which the results were published only in *Disentangling the NAEP 1985-86 Reading Anomaly* [see Yamamoto, 1990]), as well as with 1986 bridge samples. For age 17, the paced-tape sample MS-Br86 does not allow direct comparisons with a 1988 sample, but comparisons can be made with 1986 bridge samples. In 1986, the science trend items were scaled with common items from the 1977 and 1982 assessments. Because of the small number of items in common with those in the 1970 and 1973 assessments, data from those assessments were not scaled, but were linked to the trend line through mean proportion correct for common items. The 1990 trend assessments were linked to the 1970, 1973, 1977, and 1982 assessments through the 1986 assessment. Information about previous assessment years, including 1970 and 1973, is available in Chapter 11 of *Expanding the New Design: The NAEP 1985-86 Technical Report* (Yamamoto, 1988).

In addition to the samples that contributed data to the results in *Trends in Academic Progress*, two other types of trend samples were collected in 1990. One is a print sample that is comparable to an assessment of age 17 students that was part of the 1988 study of the 1986 reading anomaly. The print bridge sample, RMS-Br86, provides the link to the 1988 trend data. The booklets, presentation, time of assessment and age definition were the same as those for the 1988 science bridge sample (for which results have been published in *Disentangling the NAEP 1985-86 Reading Anomaly* only) and for part of the main NAEP sample in 1986. Only three blocks in the booklets (S1-S3) were analyzed in 1988, since the other science blocks in the booklets (S0 and S4) were there only to maintain consistency in the presentation of blocks across the 1986 main and 1988 trend assessments. Because the results from the age 17 print bridge sample were not included in *Trends in Academic Progress*, analyses of these data are not described here.

The other type of sample not contributing to the results in *Trends in Academic Progress* were administered new booklets formed with items from previous assessments. These samples include 9[MS-BrLT], 13[MS-BrLT], and 17[MS-BrLT]. They use the same season, mode, and age definitions as the samples that bridge to 1986, but differ in the booklets administered. The purpose of these samples was to supplement the number of items and students contributing to long-term trend. Future trend assessments were to include the items presented to these samples. The items in these new booklets were not previously scaled. The reasons that the data from these samples were not included in *Trends in Academic Progress* are detailed in section 14.1.2.

The numbers of scaled items for each age are presented in Table 14-3. Because of the small number of items in common across all three ages, and because no paced-tape data from 1988 were available for age 17, each age was scaled separately. Following the lead of previous trend analyses, the trend scales are univariate. Derivation of subscales for specific content areas

Table 14-2
NAEP Science Samples Contributing to 1990 Trend Results, 1970-1990

Cohort	Year	Sample	Subjects	Time of Testing	Mode of Administration	Age Definition	Modal Grade
Age 9	1970	Main	SWC	Winter	Tape	CY	4
	1973	Main	MS	Winter	Tape	CY	4
	1977	Main	Sci	Winter	Tape	CY	4
	1982	Main	MSC	Winter	Tape	CY	4
	1986	BrLT*	RMS	Winter	Mixed	CY	4
	1988	Br86*	RMS	Winter	Mixed	CY	4
	1990	Br86*	RMS	Winter	Mixed	CY	4
	1990	BrLT	MS	Winter	Tape	CY	4
Age 13	1970	Main	SWC	Fall	Tape	CY	8
	1973	Main	MS	Fall	Tape	CY	8
	1977	Main	Sci	Fall	Tape	CY	8
	1982	Main	MSC	Fall	Tape	CY	8
	1986	BrLT*	RMS	Fall	Mixed	CY	8
	1988	Br86*	RMS	Fall	Mixed	CY	8
	1990	Br86*	RMS	Fall	Mixed	CY	8
	1990	BrLT	MS	Fall	Tape	CY	8
Age 17	1970	Main	SWC	Spring	Tape	Not CY	11
	1973	Main	MS	Spring	Tape	Not CY	11
	1977	Main	SL	Spring	Tape	Not CY	11
	1982	Main	MSC	Spring	Tape	Not CY	11
	1986	Br**	MS	Spring	Tape	Not CY	11
	1990	Br86**	MS	Spring	Tape	Not CY	11
	1990	BrLT	MS	Spring	Tape	Not CY	11
Age 17/Grade 11	1986	Main*	RMSCHL	Spring	Print	Not CY	11
	1988	Br86*	RMS	Spring	Print	Not CY	11
	1990	Br86*	RMS	Spring	Print	Not CY	11

* Within a cohort, these samples received common booklets.
** These samples received common booklets. The science items in these booklets are also in booklets given to the age 17 print samples in print format.

LEGEND:

Sci	Science
MS	Mathematics and science
MSC	Mathematics, science, and citizenship
RMS	Reading, mathematics, and science
RMSCHL	Reading, mathematics, science, computer understanding, U.S. history, literature
SL	Science, life skills
SWC	Science, writing, and citizenship
Main	Main assessment
Br	Bridge assessment
Br86	Bridge to 1986
BrLT	Bridge for long-term trend
Print	Printed administration
Tape	Audiotape administration
Mixed	Mathematics and science administered by audiotape, reading administered by print
CY	Calendar year: birthdates in 1980, 1976, and 1972 for ages 9, 13, and 17
Not CY	Age 17 only: birthdates between Oct. 1, 1972 and Sept. 30, 1973

Table 14-3

Numbers of Scaled Science Trend Items Common Across Ages

Age	Booklets	Number of Items
9 only	91-93	55
13 only	91-93	30
17 only	84-85	32
9 and 13 only	91-93, 91-93	0
9 and 17 only	91-93, 84-85	0
13 and 17 only	91-93, 84-85	45
9, 13, and 17	91-93, 91-93, 84-85	1
Total		163

Table 14-4

Numbers of Scaled Science Trend Items Common Across Assessments

	Number of Items		
Assessment Years	Age 9	Age 13	Age 17
1986, 1988, 1990	56	76	78
1982, 1986, 1988, 1990	10*	58	47
1977, 1986, 1988, 1990	56	76	76
1977, 1982, 1986, 1988, 1990	10*	58	45

* Twenty-four items common to years 1977 and 1982, but not later years, were included in the 1986 scaling of these items to stabilize the estimation of the item parameters. See *Expanding the New Design: The NAEP 1985-86 Technical Report* for more information.

was not feasible given the limited number of items presented to students in the trend samples. The number of items scaled in 1990 that were common across assessment years is in Table 14-4.

The steps in the science trend analysis are documented in the following sections. As is usual in NAEP analyses, the first step was to gather item and block information. Next, the trend items were calibrated, derived background variables were calculated, and plausible values were generated after conditioning on available background variables. Finally, the scales were placed on the final science trend proficiency scale used in previous trend assessments.

14.1.1 Item Analysis for the Paced-tape Bridge-to-1986 Assessment

Conventional item analyses did not identify any difficulties with the paced-tape bridge data for the 1990 samples that bridge to 1986. Table 14-5 contains information about the science trend blocks. These blocks were presented to samples 9[RMS-Br86], 13[RMS-Br86], and 17[MS-Br86]. At ages 9 and 13, the blocks labeled S1, S2, and S3 were presented intact to 1986 and 1988 trend samples. The age 9 and age 13 blocks were in booklets 91 through 93, respectively. At age 17, S1, S2, and S3 were presented intact to the 1986 trend sample. Block S3 was in booklet 84 and blocks S1 and S2 were in booklet 85. Common labeling of these blocks across ages does not denote common items.

Table 14-5 contains the number of items, size of the sample administered the block, mean weighted proportion correct, mean weighted biserial, and mean weighted alpha as a measure of reliability for each block. The average values were calculated using examinee weights and the items in the block that were scaled. The 1990 item-level statistics were not very different from those for the 1986 and 1988 assessments. The percent of examinees not reaching items in the bridge blocks was always zero because the items were administered with a tape-recording to pace response time.

14.1.2 Item Analysis for the Paced-tape Trend Supplement

The purpose of the paced-tape bridge for long-term trend (samples 9[MS-BrLT], 13[MS-BrLT], and 17[MS-BrLT]) was to supplement the samples of items and students in the paced-tape bridge to 1986 samples. For each age, booklet 94, containing block SJ, and booklet 95, containing block SK and SL, were presented to additional samples of students with the same characteristics as those who received booklets 91-93 (blocks S1-S3) at ages 9 and 13 and booklets 84-85 (S1-S3) at age 17. The correspondence between booklets and blocks is described in Tables 14-6 and 14-7. For the 1986 assessment, blocks S1, S2, and S3 were formed using the best items available from previous assessments. For the 1990 assessment, blocks SJ, SK, and SL were formed using the best of the remaining items from previous assessments. As for S1-S3, common names for the blocks SJ through SL for the different ages does not indicate that common items were in those blocks.

In order to combine the information from the two types of samples into one scale, items from both sets of booklets needed to be similar enough to ensure that the meaning of the science trend scale would not change due to the addition of new booklets to the trend.

Table 14-5

Descriptive Statistics for Item Blocks
Science Trend Samples

Statistic	Block		
	S1	S2	S3
Age 9			
Number of scaled items	17	20	19
Number of scaled open-ended items	0	0	0
Unweighted sample size	1991	2050	2194
Average weighted proportion correct	.63	.57	.71
Average weighted r-biserial	.55	.46	.55
Weighted alpha reliability	.66	.61	.69
Age 13			
Number of scaled items	23	30	23
Number of scaled open-ended items	0	0	0
Unweighted sample size	2229	2288	2132
Average weighted proportion correct	.54	.56	.59
Average weighted r-biserial	.51	.47	.52
Weighted alpha reliability	.72	.76	.73
Age 17			
Number of scaled items	24	31	23
Number of scaled open-ended items	0	0	0
Unweighted sample size	2206	2206	2205
Average weighted proportion correct	.64	.64	.59
Average weighted r-biserial	.50	.53	.63
Weighted alpha reliability	.71	.79	.82

Table 14-6

Correspondence Between Samples, Booklets, and Blocks for Bridge-to-1986 Science Trend

Sample	Booklet	Block
Age 9 RMS-Br86	91	S1
	92	S2
	93	S3
Age 13 RMS-Br86	91	S1
	92	S2
	93	S3
Age 17 MS-Br86	84	S3
	85	S1 S2

Table 14-7

Correspondence Between Samples, Booklets, and Blocks for Supplemental Science Trend

Sample	Booklet	Block
Age 9 MS-BrLT	94	SJ
	95	SK SL
Age 13 MS-BrLT	94	SJ
	95	SK SL
Age 17 MS-BrLT	94	SJ
	95	SK SL

Preliminary item analyses (as seen in Table 14-10) indicated that the items in booklets 94 and 95 were performing differently than the items in the bridge to 1986, in a way that affected the relationship of subgroup means on the proficiency scale. At least two factors contributed to the differences between the two sets of items. First, the content characterization of the items was different across the two sets of books. Second, the number of alternatives to each of the items differed across the two groups of booklets. For these reasons, scaled results for samples 9[MS-BrLT], 13[MS-BrLT], and 17[MS-BrLT] were not included in *Trends in Academic Progress*. The report did include results of background questions on course taking at the high school level. The background questions of interest were presented only in booklet 95 to a subset of students from sample 17[MS-BrLT].

Table 14-8 lists the number of items by block and by content area as defined in *Science Objectives: 1985-86 Assessment* (NAEP, 1985b). The proportion of items in each content area differs for booklets 91-93 or 84-85 as opposed to booklets 94 and 95. The groups of booklets also differed in the number of items having two, three, four, five and six alternatives. This is documented in Table 14-9. Table 14-10 lists the mean proportion correct for each block for male and female students. A comparison of the values for the blocks presented at age 13 shows that the items in blocks SJ-SL would present a different picture of subgroup means than do blocks S1-S3. Although all the other blocks show small differences in mean proportion correct favoring male students (blocks S1 and S2 show the largest differences), block SJ has mean proportion correct values favoring female students. Because block SJ was the only science block administered to half of the supplementary sample, 13[MS-BrLT], differences in the relationship between means for male and female students for the two sets of booklets was even more striking on the proficiency scale. At age 9, block S1 has mean proportion correct values that do not differ between male and female students, while blocks SJ, SK, and SL have mean proportion correct values that consistently favor male students. At age 17, blocks SJ and SK show very little difference in mean proportion correct values for male and female students, while blocks S1, S2, and S3 show more differences between the groups. The values in Tables 14-8, 14-9, and 14-10 are based on all items within each block, rather than only those that were included in scaling.

14.1.3 Estimation of Item Parameters

The first step in the scaling process was the estimation of item parameters for the trend items. This item calibration was performed using the NAEP version (Rogers & Nelson, 1990) of the BILOG program (Mislevy & Bock, 1982) separately for each of the three age groups, using combined data from several assessment years and treating each assessment sample as a sample from a separate subpopulation. The calibration was performed on a subsample of all the available subjects, resulting in approximately 500 examinees in each assessment year for each item. For ages 9 and 13, the responses were not weighted for the initial part of the analysis, because the weights were unavailable early in the analysis process. After convergence was reached without weights, student weights were added to the analysis, and final convergence was obtained. For age 17, weights were used during the entire scaling process.

Item parameters were estimated for items in blocks S1 through S3 at ages 9 and 13 using 1986, 1988, and 1990 data with the NAEP-BILOG computer program separately for each age. Parameters were estimated for items in blocks S1 through S3 given to the age 17 paced-tape

Table 14-8
Number of Science Items by Content Area and Group of Booklets

Age	Booklets	Content Area: Life Sciences	Physical Science	Chemistry	Earth & Space Science	History of Science	Nature of Science
9	91-93	15	29	5	4	0	10
	94-95	26	17	2	17	0	2
13	91-93	15	9	10	22	1	26
	94-95	32	13	9	14	0	9
17	84-85	14	17	12	21	0	18
	94-95	26	10	12	16	0	14

Table 14-9
Number of Science Items by Number of Alternatives and Group of Booklets

Age	Booklets	Number of Alternatives: 2	3	4	5	6
9	91-93	29	10	18	6	0
	94-95	13	9	33	11	0
13	91-93	29	13	20	19	2
	94-95	20	9	28	19	1
17	84-85	38	8	17	17	2
	94-95	22	5	28	22	1

Table 14-10
Male-Female Differences in Mean Percentage Correct by Block, Science Trend

Block	Age 9			Age 13			Age 17		
	Male	Female	Difference	Male	Female	Difference	Male	Female	Difference
Regular									
S1	63	63	0	55	52	3	65	62	3
S2	55	50	5	57	54	3	67	61	6
S3	70	68	2	57	56	1	60	58	2
Supplemental									
SJ	53	50	3	56	57	-1	64	64	0
SK	62	58	4	76	74	2	58	57	1
SL	63	58	5	52	50	2	62	57	5

samples using 1986 and 1990 data. Although items were examined for irregularities, only items that were deleted from the previous scaling of the paced-tape trend data were excluded in the 1990 analysis. This was done to ensure that the equating of the 1990 scale (based on item parameters from 1986, 1988, and 1990 data) to the 1986 scale reflected the same relationships between subgroup means for both the 1986 sample as originally reported and the 1986 sample as estimated on the 1990 scale. Eleven percent of the items administered to the bridge-to-1986 sample were excluded from analyses of previous assessments. The deleted items appear in Table 14-11. As a result of these deletions, 56 items were scaled for age 9, 76 items were scaled for age 13, and 78 items were scaled for age 17. A list of the items scaled for each of the ages, along with their item parameter estimates, appears in Tables E-24, E-25, and E-26 in Appendix E.

Table 14-11
Items Deleted from the Science Paced-tape Trend Analysis

Age	Booklet	Block	Item	Reason for Exclusion
9	91	S1	N400201	Excluded in previous assessment
	92	S2	N401701	Excluded in previous assessment
	92	S2	N402003	Excluded in previous assessment
	92	S2	N402004	Excluded in previous assessment
	92	S2	N402601	Excluded in previous assessment
	92	S2	N402603	Excluded in previous assessment
	93	S3	N403802	Excluded in previous assessment
13	91	S1	N404902	Excluded in previous assessment
	91	S1	N404903	Excluded in previous assessment
	92	S2	N407501	Excluded in previous assessment
	93	S3	N409401	Excluded in previous assessment
	93	S3	N409402	Excluded in previous assessment
	93	S3	N409403	Excluded in previous assessment
	93	S3	N409801	Excluded in previous assessment
17	85	S1	N410001	Excluded in previous assessment
	85	S1	N410002	Excluded in previous assessment
	85	S1	N410301	Excluded in previous assessment
	85	S2	N407402	Excluded in previous assessment

14.1.4 Derived Background Variables

In the trend analysis, any variables derived for the science analysis from background questions were used both for conditioning and in reporting (to define subgroups). Derived conditioning and reporting variables are described in Appendix B.

14.1.5 Generation of Plausible Values

The generation of plausible values was conducted independently by age for each of the three assessment years. The item parameters from NAEP-BILOG, final student weights, item responses and selected background variables (conditioning variables) were used with the

computer program NAEP-MGROUP (Rogers, 1991; Sheehan, 1985), in order to generate the values for each age. Because there were fewer background variables available for the bridge samples, fewer conditioning variables were used in the creation of the plausible values on the trend scale than on the cross-sectional scale. There were 45 contrasts in the conditioning model at age 9, including an overall constant, 48 at age 13, and 54 at age 17. Appendix F gives the codings for the conditioning variables (Table F-7) and the estimated conditioning effects (Tables F-30, F-31, and F-32) for the three age groups. The estimated conditioning effects in the tables are expressed on the scale of the original calibration. A check on the distributions of the plausible values for each age was made.

14.1.6 The Final Proficiency Scale

The trend and cross-sectional scales were not directly linked in 1990 because the 1990 trend and cross-sectional samples were not comparable. The linear indeterminacy of the trend scale was resolved by linking the 1990 trend scales to previous trend scales. For each age, the item parameters from 1990 based on data from 1986, 1988, and 1990 were used with the 1986 data to find plausible values for the 1986 data. The mean and standard deviation of all of the plausible values were calculated and matched to the mean and standard deviation of all of the plausible values based on the 1986 item parameters and 1986 data as reported in earlier reports. The transformations that resulted from this matching of the first two moments for the 1986 data are

Age 9: $\theta_{proficiency} = 44.04 \bullet \theta_{calibrated} + 223.46$

Age 13: $\theta_{proficiency} = 38.24 \bullet \theta_{calibrated} + 254.41$

Age 17: $\theta_{proficiency} = 45.97 \bullet \theta_{calibrated} + 290.24.$

where $\theta_{proficiency}$ denotes values on the final transformed scale and $\theta_{calibrated}$ denotes values on the original calibration scale. Overall summary statistics for the trend samples are given in Table 14-12.

Table 14-12
Means and Standard Deviations on the Science Trend Proficiency Scale

		All Five Plausible Values	
Age	Assessment	Mean	S. D.
9	1986	224.3	41.6
	1990	228.7	40.2
13	1986	251.4	36.6
	1990	255.2	37.6
17	1986	288.5	44.4
	1990	290.4	46.2

14.2 CROSS-SECTIONAL DATA ANALYSIS

The data from the main focused-BIB assessment of science (9[Sci-MainP, 13[Sci-MainP], and 17[Sci-MainP]) were used for cross-sectional analyses comparing the levels of science achievement for various subgroups of the 1990 target populations. The main assessment included three student cohorts: students who were either in the fourth grade or 9 years old, students who were either in the eighth grade or 13 years old, and students who were either in the twelfth grade or 17 years old. The birth date ranges for age-eligible students were based on the 1980, 1976, and 1972 calendar years respectively for ages 9, 13 and 17. The sampled students in each of these three cohorts were assessed either in the winter or the spring. The samples in the main assessment are listed in Table 14-1.

The pool of items used in the 1990 science assessment contained a range of open-ended and multiple-choice questions measuring performance on sets of objectives documented in *Science Objectives: 1990 Assessment* (NAEP, 1989b). The objectives framework is described in Chapter 2. A total of 255 distinct science items addressing these objectives was administered in 1990 using the focused-BIB design to allocate the items to the assessed students. The items were classified into four categories based on their content: life sciences, physical sciences, earth and space sciences, and nature of science. These four categories of items determined the subscales scaled in 1990.

In these samples, each student was administered a booklet containing three blocks of science cognitive items, a block of background questions common to all booklets for a particular age/grade level, and a block of science-related background questions common to all science booklets for a particular age/grade level. Seven blocks of science cognitive questions were administered at each age/grade level in a total of seven booklets for each level. (See Chapter 4 for more information about the blocks and booklets.) Both age- and grade-selected students contributed to the cross-sectional scaling. However, the "grade-only" portion of the main focused-BIB science samples (whether the time of assessment was winter or spring) contributed to the means and percentages of the cross-sectional results that are reported in *The 1990 Science Report Card.*

It should be noted that, although open-ended items were scattered throughout the seven cognitive blocks for each level, one block (block SH) consisted of only open-ended figural-response items. As were the other open-ended items, these items were scored dichotomously and were scaled with the multiple-choice items, but they were also analyzed independently.

In the final scale, there were 217 multiple-choice items and 27 open-ended items, for a total of 244 items. Three of these items were treated as separate items for different grades (see Table 14-17). The number of overlapping items for the age/grade levels are listed in Table 14-13. Numbers of items in each subscale by block and by booklet are given in Tables E-6 and E-7 in Appendix E.

Table 14-13
Number of Scaled Science Items Common Across Grades, by Subscale

Grade	Life Sciences	Physical Sciences	Earth & Space Sciences	Nature of Science	Total
4 only	8	12	9	6	35
8 only	7	5	6	6	24
12 only	18	28	12	9	67
4 and 8 only	12[1]	14	12	5	43
8 and 12 only	13[2]	11	13	6	43
4 and 12 only	0	0	0	0	0
4, 8, and 12	10	9[3]	10	3[4]	32
Total	68	79	62	35	244

[1]One of these items was treated as different items for grade 4 and grade 8.

[2]An additional unscaled open-ended item (K027401) was administered to grade 8 and grade 12 students and was scored in two ways. The second score for the item (labeled K027402) was not analyzed, even in initial item analyses.

[3]One of these items was dropped for grade 4: another was treated as different items for grade 4 and for grades 8 and 12.

[4]One of these items was treated as different items for grade 4 and for grades 8 and 12.

The next sections contain in some detail a description of the analysis performed using the main focused-BIB sample data. As for the trend analysis, the process began with an examination of the items and blocks of items. Open-ended items were dichotomously scored, items were grouped by content domains, and derived background variables were calculated. The estimation of item parameters for the four science subscales was completed, followed by the generation of multivariate plausible values. Finally, the plausible values were transformed to the final proficiency scale.

14.2.1 Item Analysis

Tables 14-14, 14-15, and 14-16 show the number of items, mean proportion correct, mean r-biserial, and alpha reliability for each block administered at each age/grade level. These values were calculated for the dichotomously scored multiple-choice and open-ended items within a block, if they were used in the scaling process. The table also gives the number of students who were administered the block and the percent not reaching the last item in the block. These numbers only include the students in the grade-only portion of the samples that contributed to the summary statistics provided in *The 1990 Science Report Card*. Student weights were used for all statistics, except for the sample sizes. The results for the blocks administered to each age/grade level indicated that the blocks differed in number of items, average difficulty, reliability, and percent not reaching the last item, and so are not parallel to one another. In particular, the figural-response items in block SH tended to be more difficult than the items in other blocks, especially for grade 4 students. Preliminary item analyses for all items within a

Table 14-14

Descriptive Statistics for Item Blocks by Position Within Booklet and Over All Occurrences
Science Cross-sectional Sample, Grade 4

Statistic	Block Position	Block SC	SD	SE	SF	SG	SH	SI
Number of scaled items	—	17	20	19	12	20	10	12
Number of scaled open-ended items	—	0	0	0	3	0	10	3
Unweighted sample size	1	932	914	897	897	888	881	878
	2	885	930	912	893	894	879	872
	3	878	873	876	918	898	858	883
	ALL	2695	2717	2685	2708	2680	2618	2633
Average weighted proportion correct	1	.52	.41	.44	.46	.57	.25	.40
	2	.51	.41	.44	.45	.56	.24	.42
	3	.51	.40	.42	.45	.55	.22	.37
	ALL	.51	.41	.43	.45	.56	.24	.40
Average weighted r-biserial	1	.53	.44	.45	.54	.52	.64	.52
	2	.51	.46	.48	.55	.50	.61	.53
	3	.53	.43	.47	.55	.56	.66	.52
	ALL	.52	.44	.47	.55	.53	.64	.52
Weighted alpha reliability	1	.68	.65	.64	.56	.70	.56	.52
	2	.66	.66	.69	.62	.68	.51	.59
	3	.69	.63	.66	.58	.76	.61	.51
	All	.68	.65	.66	.58	.72	.56	.55
Weighted proportion of students attempting last item	1	.97	.82	.86	.76	.93	.79	.73
	2	.94	.86	.90	.78	.89	.83	.73
	3	.96	.83	.90	.84	.94	.68	.76
	ALL	.96	.84	.89	.79	.92	.77	.74

Table 14-15

Descriptive Statistics for Item Blocks by Position Within Booklet and Over All Occurrences
Science Cross-sectional Sample, Grade 8

Statistic	Block Position	Block SC	SD	SE	SF	SG	SH	SI
Number of scaled items	—	22	26	25	17	22	14	16
Number of scaled open-ended items	—	0	0	0	3	2	14	3
Unweighted sample size	1	904	926	934	940	933	933	946
	2	946	903	926	933	941	929	945
	3	930	942	944	901	921	923	938
	ALL	2780	2771	2804	2774	2795	2785	2829
Average weighted proportion correct	1	.50	.57	.64	.62	.43	.41	.60
	2	.51	.55	.62	.63	.42	.40	.61
	3	.50	.55	.60	.60	.40	.40	.58
	ALL	.50	.56	.62	.62	.42	.40	.60
Average weighted r-biserial	1	.50	.51	.57	.51	.49	.57	.54
	2	.50	.53	.57	.54	.46	.62	.57
	3	.53	.54	.57	.53	.45	.61	.57
	ALL	.51	.53	.57	.53	.47	.60	.56
Weighted alpha reliability	1	.74	.78	.82	.66	.73	.66	.62
	2	.74	.80	.83	.69	.69	.71	.67
	3	.78	.81	.82	.69	.65	.73	.68
	ALL	.75	.80	.82	.68	.69	.70	.66
Weighted proportion of students attempting last item	1	.91	.92	.91	.93	.58	.90	.99
	2	.91	.94	.90	.92	.61	.89	.98
	3	.93	.93	.94	.91	.70	.81	.98
	ALL	.92	.93	.92	.92	.63	.87	.98

Table 14-16

Descriptive Statistics for Item Blocks by Position Within Booklet and Over All Occurrences
Science Cross-sectional Sample, Grade 12

Statistic	Block Position	SC	SD	SE	SF	SG	SH	SI
Number of scaled items	—	25	29	21	16	25	16	10
Number of scaled open-ended items	—	0	0	0	1	2	16	2
Unweighted sample size	1	905	908	911	909	922	885	875
	2	874	903	908	909	907	916	892
	3	918	892	873	901	907	899	902
	ALL	2697	2703	2692	2719	2736	2700	2669
Average weighted proportion correct	1	.61	.64	.48	.44	.52	.47	.53
	2	.61	.66	.49	.44	.52	.46	.54
	3	.59	.65	.46	.45	.51	.46	.50
	ALL	.60	.65	.48	.44	.52	.46	.52
Average weighted r-biserial	1	.54	.57	.50	.42	.53	.64	.57
	2	.53	.58	.55	.43	.53	.64	.60
	3	.56	.58	.54	.45	.54	.62	.62
	ALL	.54	.58	.53	.43	.53	.63	.60
Weighted alpha reliability	1	.79	.83	.72	.45	.79	.79	.53
	2	.79	.84	.78	.48	.78	.78	.58
	3	.82	.85	.76	.52	.80	.76	.62
	ALL	.80	.84	.76	.49	.79	.77	.58
Weighted proportion of students attempting last item	1	.93	.92	.82	.70	.52	.54	.75
	2	.87	.92	.86	.72	.52	.57	.73
	3	.92	.93	.87	.75	.66	.53	.64
	ALL	.90	.92	.85	.73	.57	.55	.71

block were completed before scaling; however, the results shown here indicate the characteristics of the items that contributed to the final scale.

Tables 14-14 through 14-16 also contain information about the effect of the position of blocks within booklets on the average percent correct for items within each block presented to the focused-BIB samples for each grade. The averages for the grade-only portion of the focused-BIB samples show that the order of blocks within booklets did not have a large or consistent effect on proficiency in the science focused-BIB assessment. Verification of the appropriate classification of the iten s in each of the four subscales (life sciences, physical sciences, earth and space sciences, and nature of science) was completed. The items with their classifications are listed in the data appendix of *The 1990 Science Report Card*. The items are also listed by subscale in Tables E-27 through E 30 in Appendix E of this document.

14.2.2 Scoring the Open-ended Items

As indicated earlier, the science assessment included many open-ended items. These items were included in the scaling process, but were also analyzed separately. Chapter 7 contains the means and ranges for percent agreement between raters for the items as they were originally scored. The right/wrong scoring of the categories of responses for the items are indicated in Tables G-5, G-6, and G-7 in Appendix G. The tables also show the percent agreement for the raters, reliability, and Cohen's Kappa, calculated after the items were dichotomized. The sample sizes listed in these tables represent the size of the samples used in calculating the rater reliability information; that is, the number of responses that were rescored for each item.

14.2.3 Estimation of Item Parameters

The NAEP-BILOG computer program (Rogers & Nelson, 1990; Mislevy & Bock, 1982) was used to estimate the item parameters of the three-parameter IRT model for the 244 items (plus 3 items that were treated as different items for different grades), using a random subsample of 12,776 (weighted N was 12,831.5) of the 25,572 (weighted N was 25,601.7) students in the main assessment samples. Items presented to the three age/grade groups were calibrated together, allowing for different means and variances across grades, using NAEP-BILOG in a mode that calculates parameters for the four subscales in the same computer run. All items had at least 1,000 responses in the subsample. The actual range of weighted responses per item was from 1,088 to 5,474. After examination of the item response functions (IRFs) of the items for differential item functioning across groups, students from all three age/grade groups and both the winter and spring administrations were included in the scaling process. The responses of students were weighted. (See Chapter 11 for further descriptions of the scaling process.)

Fewer than six percent of the items were deleted from the IRT analysis or treated as different items for separate groups due to lack of fit to the model. These items are listed in Table 14-17, along with the reason for special treatment. In some blocks, small groups of items

Table 14-17

Items from the Science Cross-sectional Analysis Receiving Special Treatment

Item	Grade	Block	Subscale	Treatment	Reason for Treatment
K010201	4, 8, 12	SD	Physical Sciences	Dropped from grade 4	Nonmonotonicity, differing IRFs
K011901	4, 8, 12	SD	Nature of Science	Separate estimates for grade 4	Differing IRFs
K013701	4, 8	SE	Life Sciences	Separate estimates for grade 4	Differing IRFs
K015601	8	SE	Physical Sciences	Dropped from grade 8	Nonmonotonicity
K016801	12	SE	Earth and Space Sciences	Dropped from grade 12	Nonmonotonicity
K017601	12	SE	Physical Sciences	Dropped from grade 12	Nonmonotonicity
K030801	12	SE	Physical Sciences	Dropped from grade 12	Nonmonotonicity
K018801	4, 8	SF	Physical Sciences	Dropped from grades 4 and 8	Dependency
K021101	12	SF	Physical Sciences	Dropped from grade 12	Nonmonotonicity
K025501	8, 12	SG	Physical Sciences	Dropped from grades 8 and 12	Dependency
K026301	4, 8, 12	SH	Physical Sciences	Separate estimates for grade 4	Differing IRFs
K029301	8	SI	Earth and Space Sciences	Dropped from grade 8	Nonmonotonicity, differing IRFs
K029901	12	SI	Physical Sciences	Dropped from grade 12	Nonmonotonicity
K030601	12	SI	Earth and Space Sciences	Dropped from grade 12	Dependency
K030603	12	SI	Earth and Space Sciences	Dropped from grade 12	Dependency

were related to one another in a way that required correct responses for certain items in order to respond correctly to a later item. In a group of dependent items, all but one of the items were dropped from the scaling, although information for all items was examined independently from the scaling. If IRFs for items presented to the various age/grades differed for some of the groups, the item was treated as different items for the groups. If items had IRFs that were nonmonotonic, the items were deleted from scaling. No items had different IRFs for the two administration times (winter and spring) for any age/grade group.

Tables E-27 through E-30 in Appendix E list the estimated item parameters. These item parameter estimates are direct output from the NAEP-BILOG program, prior to any rescaling and specifying six subpopulations (each age/grade for the winter administration and each age/grade for the spring administration). Because of the indeterminacy of the IRT scale, the origin and size of the scale were set provisionally by standardizing the distribution of the calibration sample of examinees to have a mean of zero and a variance of one.

14.2.4 Derived Background Variables

Derived variables based upon background questions were used for two purposes: as conditioning variables, or as reporting variables used to define subgroups. Some of these variables are common to all the subject areas; others are specific to the 1990 science assessment. Derived conditioning and reporting variables are described in Appendix B.

14.2.5 Generation of Plausible Values and Calculation of Composite Plausible Values

Multivariate plausible values were generated for each age/grade group separately using the NAEP-MGROUP program. Final student weights were used at this stage of the analysis. Instead of using selected background variables as conditioning variables in this process, principal components of those variables explaining 90 percent of the variance contained in them were used as conditioning variables.

The codings of the original science-specific conditioning variables, before principal components were calculated, are presented in Appendix F in Table F-6. Common core conditioning variables were also used. (For age 17/grade 12, the "modal age, > modal grade" category was deleted from the age-by-grade variable, because students above grade 12 were not sampled.) The proportion of variance of each original conditioning variable accounted for by the principal components included in the conditioning model is listed in Tables F-24, F-26, and F-28 in Appendix F.

The estimated conditioning effects for the principal components of the three samples defined by the three age/grade groups are given in Appendix F in Tables F-25, F-26, and F-27. The values of the conditioning effects are expressed in the metric of the original calibration scale. Definitions of derived conditioning variables are given in Appendix B.

14.2.6 The Final Proficiency Scale

In order to resolve the linear indeterminacy of the original IRT calibration scale, the proficiency results from the main focused-BIB samples were tied to the 1986 cross-sectional results. Although the science proficiency scale, because of this selection, is seemingly expressed in the same units as those of other NAEP proficiency scales, it is not appropriate to compare science proficiency scores with scores on the other subject area scales. Any other convenient transformation of the original science plausible values could have been chosen, so there is no link in the construction of the science proficiency scale to the scales of any other subject area.

The connection between the 1990 cross-sectional and 1986 scales is weak because of the change in target samples specified in the 1989 NAEP legislation and because of new items representing a new viewpoint of the 1990 science objectives (see Chapter 2 for more information about the objectives). The 1990 main-BIB samples have calendar-year age definitions with the modal grades 4, 8, and 12, and the assessment took place in the winter and spring. In 1986, the main-BIB samples had age definitions based on an October 1 to September 30 year with the modal grades 3, 7, and 11. The assessment took place in the spring only. In addition, the booklets administered to the main-BIB samples in 1990 were focused-BIB spiraled, containing only background blocks and three science cognitive blocks. The main-BIB samples in 1986 were administered booklets with subject-area cognitive blocks that were spiraled in a balanced incomplete block design, so that some examinees received only one science cognitive block.

A complete 1990 bridge back to the 1986 cross-sectional assessment of reading, mathematics, computer competence, and science at grades 3, 7, and 11 would have been prohibitively expensive. Alternatively, a complete bridge from the 1990 main NAEP samples to the current 1990 bridge to 1986 could have been implemented, at great cost, as was done to bridge the 1986 main NAEP samples to the 1986 bridge samples to the past. (The 1986 9[RMS-BrLT], 13[RMS-BrLT], and 17[MS-Br] samples in Table 14-2 and the 1986 9[MS-Br] and 13[MS-Br] samples not listed in Table 14-2 fulfill this purpose.) This lack of comparable samples across years is similar to the situation encountered in trying to linking the 1988 reading cross-sectional scale to the 1984 reading scale. After attempting three links between the two reading scales and finding disparate results, the 1988 reading cross-sectional scale was not linked to the 1984 scale. However, a plan for linking the 1986 and 1990 science scales in some way was requested by the National Center for Education Statistics.

For the 1990 cross-sectional assessment, seven science cognitive blocks were presented in focused-BIB booklets at each age/grade. Six of the seven blocks for each age/grade contain only new items written to fit the 1990 objectives. For each age/grade, the seventh block (block SC) contains items in print form that also appear in one of three paced-tape bridge blocks that were administered in 1986, in 1988 (print only for age 17) and in 1990. These three blocks were also administered in print form for the 1986 cross-sectional assessment. Of the 49 items presented in the 1990 blocks containing old items, an average of 21 items are available for each age/grade sample with an overlap of 13 items for the age 13/grade 8 and age 17/grade 12 samples. The upper grade samples share only 2 items with the age 9/grade 4 sample.

There are four possible ways to connect the 1990 cross-sectional scale with the 1986 scale. Three of these are based on connections between the 1990 cross-sectional and trend

samples. One is based on a direct connection between the 1990 and 1986 cross-sectional scales using the common items described above.

The connections between the 1990 cross-sectional and trend scales include commonly defined subsamples from the following samples (see Tables 14-1 and 14-2):

1) the spring portion of 1990 17[Sci-MainP] and 1986 17[MS-Br] (tape);
2) the spring portion of 1990 17[Sci-MainP] and 1988 17[RMS-Br86] (print); and
3) the winter portion of 1990 9[Sci-MainP] and 1986 9[RMS-BrLT] (tape).

In the past, when connecting main NAEP and bridge samples for an assessment using IRT methods, there have been three blocks of items that differed only in the mode of administration (print or tape) presented to each of the samples, the minimum sample size in an age/grade sample was about 3,800, and links were available at all three ages. For options 1 and 2, there would be 25 common items, only at age 17. For both of these options the items differ in context, and, for option 1, they also differ in mode of administration. Using the common age definition of October 1 through December 30, 1972 for the commonly defined subsamples for options 1 and 2 will limit the sample size for the main NAEP subsample to less than 429 (3/7 of 1,000). Option 2 was not seriously considered, because in addition to these problems, it depends on the link of the 1988 science bridge data to the 1986 scale. For option 3, only 17 items are in common for the two samples, and the connection can only be made at age 9. The sample size for the main NAEP commonly defined subsample would be about 1,286 (3/7 of 3,000).

Because of these limitations, the alternative of linking directly through the items common to both the 1986 and 1990 main-BIB samples was selected. Although the linking items are limited to the 49 items described above, 3,597, 3,738, and 3,606 students from the 1990 main-BIB samples for age 9/grade 4, age 13/grade 8 and age 17/grade 12, respectively, and 6,978, 7,131, and 6,974 examinees from the 1986 main BIB samples at each age/grade took some of the common items and, so, were available to tie the scales together.

The appropriate transformations were found separately for age 9/grade 4 and for age 13/grade 8 and age 17/grade 12. This was done because of the small number of items presented at age 9/grade 4 in common with items presented at the other ages. The processes were parallel for the two links, so only the age 9/grade 4 link will be described.

First, information from both the 1986 and 1990 assessments was pooled by scaling the items in the 1990 common item block and the items in the three 1986 blocks containing those items using 1990 age 9/grade 4 students receiving the common item block and 1986 age 9/grade 3 students receiving the three blocks. This produced new item parameter estimates for the linking items. Because the linking items did not appear as an intact block of items in the 1986 assessment, there was the possibility that at least some of the items performed differently in the two years due to context and position effects. The fit of the item parameter estimates to the 1990 and the 1986 data was evaluated, but no items performed differently in the two assessments, so no items were removed from the linking for this reason. The distributions of proficiencies based on all five plausible values for each of the 1990 and 1986 linking groups were found. This was done by using the full set of principal components of background variables in the conditioning model for the 1990 group. However, for the 1986 linking group only the overall conditioning constant was used in the conditioning model.

Next, using the 1986 item parameters for the items in the three 1986 blocks, the distribution of scaled proficiencies was found for the 1986 linking group using the same conditioning model containing only the overall constant. The transformation matching the mean and standard deviation of all five plausible values for the 1986 group of students based on the new item parameters to the mean and standard deviation of all five plausible values for the 1986 group of students based on the 1986 item parameters was calculated. This transformation was used to place the 1990 students receiving the common item block on a scale related to the 1986 scale. Finally, the 1990 distribution of the science composite for age 9/grade 4 was transformed so that the mean and standard deviation of all five plausible values have the same mean and standard deviation as all five plausible values for the 1990 age 9/grade 4 students who received the common block.

The transformations from the original calibration scale to the final proficiency scale are

Age 9/Grade 4: $\theta_{proficiency} = 49.39 \bullet \theta_{calibrated} + 267.22$, and

Age 13/Grade 8
and Age 17/Grade 12: $\theta_{proficiency} = 64.32 \bullet \theta_{calibrated} + 246.36$.

These transformations were used for each subscale, as well as for the composite. Overall summary statistics for the samples are presented in Table 14-18. Direct comparison of the 1986 and 1990 cross-sectional results are limited by the different age definitions and times of testing for the two assessments, as well as the frailty of this 1986-1990 linking process.

Table 14-18
Means and Standard Deviations on the Science Cross-sectional Composite Proficiency Scale

Grade	All Five Plausible Values	
	Mean	S. D.
4	232.8	31.3
8	263.1	39.5
12	293.3	42.5

The composite scale is defined as a weighted average of the results across subscales, where the weights differ by age/grade and are assigned to be proportional to the percentage distribution of items by content. The proportions of items from each subscale within the science cross-sectional assessment were specified in *Science Objectives, 1990 Assessment* (NAEP, 1989b) and are repeated here in Table 14-19.

Table 14-19
Defining Weights for the Science Composite by Age/Grade

Subscale	Age 9/Grade 4	Age 13/Grade 8	Age 17/Grade 12
Life Sciences	30	30	32
Physical Sciences	30	30	34
Earth and Space Sciences	30	30	22
Nature of Science	10	10	12

14.2.7 Partitioning of the Estimation Error Variance

The variance of subscale and composite means for each grade was partitioned into the part due to the sampling of students and the part due to the latency of proficiency, θ, as described in Chapter 11. Table 14-20 contains estimates of the sampling variance, U*, and a multiple of the estimates of the variance among the weighted means of the five plausible values transformed to the final proficiency scale. The table also contains an estimate of the total error variance, V, as well as the proportion of error variance due to sampling students and due to the latent nature of θ.

Table 14-20
Estimation Error Variance and Related Coefficients for the Science Cross-sectional Assessment

Grade	Scale	U*	$(1+5^{-1})B$	V	Proportion of Variance Due to... Student Sampling: U*/V	Latency of θ: $(1+5^{-1})B/V$
4	Composite	0.81	0.01	0.82	0.98	0.02
	Life sciences	0.81	0.04	0.85	0.95	0.05
	Physical sciences	1.06	0.13	1.19	0.89	0.11
	Earth & space sciences	0.80	0.09	0.88	0.90	0.10
	Nature of science	0.85	0.14	0.98	0.86	0.14
8	Composite	1.46	0.03	1.50	0.98	0.02
	Life sciences	1.44	0.05	1.48	0.97	0.03
	Physical sciences	1.48	0.03	1.51	0.98	0.02
	Earth & space sciences	1.74	0.10	1.84	0.94	0.06
	Nature of science	1.71	0.22	1.93	0.88	0.12
12	Composite	1.42	0.03	1.45	0.98	0.02
	Life sciences	1.08	0.09	1.16	0.93	0.08
	Physical sciences	2.08	0.16	2.25	0.93	0.07
	Earth & space sciences	1.50	0.19	1.69	0.89	0.11
	Nature of science	1.60	0.12	1.72	0.93	0.07

14.2.8 Science Teacher Questionnaire

The teacher survey included those teachers who were identified as teaching eighth-grade students who were assessed in science. Variables derived from the questionnaire were used in the conditioning model for the age 13/grade 8 sample, along with a variable that indicated whether a student record was matched or partially matched by a teacher record, so that means for subgroups defined by these variables could be compared with no bias. Of the 575 questionnaires that were distributed, 544 were returned by teachers; of those, 510 matched at least one student record. Of the 6,531 eighth-grade students, 4,799 were matched with both parts of the teacher questionnaire, for a match rate of 73.5 percent, and 787 were matched with only the first part of the teacher questionnaire, for a match rate of 12.1 percent. Thus, 85.6 percent of the students were matched with at least the background information about their science teachers.

14.2.9 Mantel-Haenszel DIF Analyses

Differential item functioning (DIF) analyses of the main focused-BIB science items used in scaling were completed after scaling to provide information to test developers for developing future science assessments. Sample sizes were large enough to compare male and female students, White and Black students, and White and Hispanic students using the Mantel-Haenszel procedure described in Chapter 9 to identify items that should be examined more closely ("C" items). DIF analyses were conducted separately by grade. A given item was subjected to at least three, and as many as nine, separate DIF analyses. Table 14-21 summarizes information about the identified items for each block.

Thirty-four of the 244 scaled cross-sectional science items were categorized as "C" items in one or more of the DIF analyses. Table H-3 (in Appendix H) identifies items that were categorized as "C" items in at least one analysis. The block containing the item, the grade, and the analysis for which the item was identified are also given in Table H-3. Contrary to experiences with other tests, if an item was flagged as having DIF in the comparison of male and female students, the item was most likely to favor the female students. In addition, only two items in block SH, the block containing only open-ended items, were identified as being "C" items. Only two items were found to have possible bias by a committee of test and subject-area experts convened to examine the "C" items. One of these items, K011101, has a stem that could mislead Hispanic students; while the other, K012201, contains gender-specific alternatives. Neither of these items will be retained for future science assessments.

14.2.10 Analysis of Dimensionality

After scaling, the relationships between the subscales were examined to provide insight into the meaning of the scales. This was done using confirmatory factor analysis on testlets formed within each booklet at each grade. The seven booklets at each grade support seven replications of the analysis. For each booklet at each age, the items were divided into groups based on subscale membership. Because of the small number of items in the fourth subscale, the nature of science, this subscale was deleted from this analysis. Then each of the four groups

Table 14-21

Numbers of "C" Items Favoring Each Group by Science Block

Block	Grade	Number of Scaled Items in Block	Analysis					
			Male/Female		White/Black		White/Hispanic	
			Male	Female	White	Black	White	Hispanic
SC	4	17	0	2	0	0	0	0
	8	22	0	2	0	0	0	0
	12	25	2	4	0	0	0	0
SD	4	20	0	0	0	0	1	0
	8	26	0	2	0	0	0	0
	12	29	1	4	1	0	0	0
SE	4	19	0	0	0	0	0	0
	8	25	1	0	0	2	0	0
	12	21	1	1	0	0	0	0
SF	4	12	0	1	1	0	0	0
	8	17	0	1	1	0	0	0
	12	16	0	0	1	0	0	0
SG	4	20	0	1	0	1	0	0
	8	22	0	0	0	0	0	0
	12	25	1	0	0	1	0	0
SH	4	10	0	1	0	0	0	0
	8	14	0	1	0	0	0	0
	12	16	0	0	0	0	0	0
SI	4	12	1	0	0	0	0	0
	8	16	3	0	1	0	0	0
	12	10	0	1	0	0	0	0

of items were randomly divided in half, so that six groups of items were identified for each booklet. For each student who received a certain booklet, the number correct was calculated for each of the six groups of items. These six scores for each student were analyzed using a factor analytic model that posited that the pairs of groups of items that were from the same subscale loaded on the same factor. The correlations between the three subscale factors in this model, averaged over the seven booklets for each age, are shown in Table 14-22.

Table 14-22
Average Correlations Between Factors

Grade	Factor	Life Sciences	Physical Sciences	Earth & Space Sciences
4	Life Sciences	1.00		
	Physical Sciences	.94	1.00	
	Earth & Space Sciences	.96	.94	1.00
8	Life Sciences	1.00		
	Physical Sciences	.96	1.00	
	Earth & Space Sciences	.94	.96	1.00
12	Life Sciences	1.00		
	Physical Sciences	.91	1.00	
	Earth & Space Sciences	.94	.96	1.00

14.2.11 Anchoring the Points on the Science Proficiency Scale

The main NAEP science composite scale was anchored in 1986, using the process described in *Expanding the New Design: The 1985-86 Technical Report*. Because each of the 1990 scales was tied to the 1986 cross-sectional or trend scale, the distribution of proficiency scores derived from the main and bridge samples can be described in terms of scale anchors. In 1986 the levels of science proficiency were

150 - Knows everyday science facts;
200 - Understands simple scientific principles;
250 - Applies basic scientific information;
300 - Analyzes scientific procedures and data; and
350 - Integrates specialized scientific information.

Chapter 15

DATA ANALYSIS FOR THE WRITING ASSESSMENT[1]

Angela M. Grima and Eugene G. Johnson

Educational Testing Service

This chapter describes the analyses carried out on the responses to the writing tasks and the background items in the 1984, 1988, and 1990 assessments of writing. These analyses led to the results reported in *Trends in Academic Progress: Achievement of U.S. Students in Science, 1969-70 to 1990; Mathematics, 1973 to 1990; Reading, 1971 to 1990; and Writing, 1984 to 1990* (Mullis, Dossey, Foertsch, Jones, & Gentile, 1991) and in *The Writing Students Do in School: The 1990 NAEP Portfolio Study of Fourth and Eighth Graders' School-based Writing* (Gentile, 1992). The emphasis is on the methods and results of the procedures used to develop the composite scores (e.g., average response method and meanparts) that formed the basis of those reports.

The objectives of the 1990 writing analyses were to:

- measure trends in writing achievement over the years 1984, 1988, and 1990; and
- analyze the data from the writing portfolio study.

Trends in writing achievement were measured by comparing the responses to a set of writing tasks by students assessed in 1990 with the responses to the same set of writing tasks for students assessed in 1988 and in 1984. The major analyses were made for trends in average task accomplishment (primary trait), although trends in writing mechanics and trends in overall writing fluency (based on holistic scoring) were also measured. The data forming the basis for these analyses are defined below. The techniques used to measure trends in writing achievement are discussed in section 15.1.

The writing portfolio study, a pilot study, consisted of random samples of fourth- and eighth-grade students who took English/language arts courses, and who participated in the 1990 writing trend assessment. The teachers of these students were asked to provide one piece of writing prepared by the student in response to a class assignment in their class. The objective of the study was to investigate the type of writing that fourth- and eighth-grade students are doing as part of English/language arts instruction, and examine the relationships between the

[1]The statistical programming for the methods (e.g., average response method and meanparts) used was performed by Bruce Kaplan and Michael Narcowich. Data analysis and additional statistical programming (e.g., almanacs, writing portfolio) were performed by Bruce Kaplan, Lucie Chan, Yim Fai Fong, Michael Narcowich, and Inge Novatkoski. Linda Lelie and Judith Alfort helped with the preparation of the tables.

characteristics and quality of students' school-based writing and their writing behaviors and achievement. Section 15.2 provides a description of the analyses conducted on the writing portfolio data.

The specific samples used for the analysis of writing achievement in 1984, 1988, and 1990 by age cohort are presented in Table 15-1.

Table 15-1
NAEP Writing Student Samples, 1984-1990

Sample	Booklets	Mode	Cohort Assessed	Time of Testing	Age Definition	Modal Grade	Number Assessed
84: 9 [RW-Main]	1-63	Print	Age 9/grade 4	Winter	CY	4	26,087
84:13 [RW-Main]	1-63	Print	Age 13/grade 8	Fall	CY	8	28,405
84:17 [RW-Main]	1-63	Print	Age 17/grade 11	Spring	Not CY	11	28,861
88: 9 [RW-Br84]	51-56	Print	Age 9/grade 4	Winter	CY	4	5,188
88:13 [RW-Br84]	51-56	Print	Age 13/grade 8	Fall	CY	8	5,500
88:17 [RW-Br84]	51-56	Print	Age 17/grade 11	Spring	Not CY	11	4,622
90: 9 [RW-Br84]	51-56	Print	Age 9/grade 4	Winter	CY	4	5,926
90:13 [RW-Br84]	51-56	Print	Age 13/grade 8	Fall	CY	8	6,233
90:17 [RW-Br84]	51-56	Print	Age 17/grade 11	Spring	Not CY	11	5,614

LEGEND:

RW	Reading and writing	CY	Calendar year: birthdates in 1980 and 1976 for ages 9 and 13.
Main	Main assessment		
Br84	Bridge to 1984	Not CY	Age 17 only: birthdates between Oct. 1 and Sept. 30.
Print	Printed administration		

15.1 TREND DATA ANALYSIS

The data contributing to the 1990 trend points for writing comes from the 1990 bridge-to-1984 samples, which match the 1984 assessments in terms of the time of administration and age definitions. The 1988 trend points also come from samples with the same characteristics as the 1984 assessment. A description of the procedures used in linking 1988 to 1984 data is found in *Focusing the New Design: The NAEP 1988 Technical Report* (Johnson & Zwick, 1990). All analyses of trends in writing performance were based on grade-eligible students only. For reasons given below, the 1990 point was determined by scores provided by raters scoring the papers in 1990. However, both the 1984 *and* the 1988 points were determined by scores provided by raters scoring the papers in 1988.

The items on which the trends in writing achievement are based are shown in Table 15-2. The table shows the block that contained the item in 1984 and trend booklets containing the item in 1988 and 1990. Twelve writing tasks were used to measure trends, with six tasks presented at each grade in 1984, 1988, and 1990. To allow comparisons in writing ability across grades, three of the six tasks presented to fourth-grade students in 1990 were also presented to eighth-grade students; three of the eighth-grade tasks were also presented to eleventh-grade students; one of the common tasks was presented at all three grades.

Table 15-2
Assignment of 1984-1990 Writing Trend Items in 1984, 1988, and 1990

Writing Task	1984 BIB-spiral Blocks			1988 and 1990 Bridge-to-1984 Booklets		
	Grade 4	Grade 8	Grade 11	Grade 4	Grade 8	Grade 11
N0003 Recreation Opportunity	-	C	C	-	52,54*	52,54
N0004 Food On Frontier	D	D	D	-	51,54*	51,54
N0005 Dissecting Frogs	-	E	-	-	53,55*	-
N0006 XYZ Company	E	E	-	52, 54*	53,55*	-
N0009 Radio Station	G	G	-	54*,55*	55*,56	-
N0010 Appleby House	G	G	G	54*,55*	55*,56	55,56
N0076 Flashlight	V	V	V**	56	-	-
N0147 Plants	C	-	-	51,53	-	-
N0148 Spaceship	E	-	-	52,54*	-	-
N0180 Space Program	-	-	E	-	-	53,55
N0190 Job Application	-	-	E	-	-	53,55
N0210 Bike Lane	-	-	G	-	-	55,56

* Only fourth- and eighth-grade students who were administered either booklet 54 or 55 were asked to participate in the writing portfolio study.

** Block V never appeared with any other writing block in 1984 (all other blocks appeared with every other block at the same grade in 1984).

15.1.1 Primary Trait Scoring of the Writing Tasks and Measures of Scorer Effect

All writing exercises from the 1990 assessment were scored for task accomplishment (primary trait). For the purposes of analysis, the student responses were coded as 0=not rated, 1=unsatisfactory, 2=minimal, 3=adequate, and 4=elaborated. "Not-reached" items were excluded from the analysis. (The writing trend blocks contained either one or two cognitive items. If an item was left blank in a one-item block, the item was scored as an omission. Items considered not-reached occurred only in writing blocks that had two cognitive items.) A 25 percent random subsample of all 1990 papers scored were rescored by a second rater to provide an estimate of interrater reliability.

Although the measures of scorer agreement in NAEP have been consistently high, we recognized the possibility that there might be variation between the ratings provided by the group of scorers assembled in 1990 and the scorers assembled in 1988. If present, this variation would add a confounding effect in the measurement of trend. The most direct way of controlling the effect of across-year variation in scoring would be to eliminate it entirely by rescoring all of the 1988 data, using the same set of scorers who scored the 1990 data. Unfortunately, resources did not allow for the rescoring of the full set of 1988 writing papers but did allow for a rescoring of approximately 7,400 of the papers given in 1988. The rescored papers for a given item constituted approximately a 25 percent sample of all 1988 papers and

consisted of all grade-eligible respondents to two or three of the 1988 booklets containing that item.

The hope was that the between-year variability in scoring would be low enough to permit the use of the full set of the 1988 data. Table 15-3 shows scorer reliability for each essay as measured by the intraclass correlation for 1988 and 1990 data by the 1988 and 1990 raters, respectively; the percentage of exact agreement between first and second raters is also given. In addition, it shows the intraclass correlation and percentage of exact score agreement comparing the scores of the 1990 raters with those of the 1988 raters on a sample of the 1988 papers. Although the reliabilities and percents of exact agreement (between first and second raters) were generally high for 1990 data, they were somewhat lower than those obtained for prior assessments (e.g., 1984 and 1988). However, the results also suggest that the variability of scoring between years was similar to the variability in scoring within the 1990 assessment.

In order to deal with the lower level of raters' consistency at the 1990 scoring, some statistical adjustments were considered. One adjustment involved removing, when possible, the scores assigned by raters that were scoring either consistently higher or consistently lower than the other raters. The other adjustment was an application of a technique developed by Braun (1988), which estimates the reliability after adjusting for systematic rater effects. Because the results obtained by these methods indicate that hardly anything would be gained by applying either of them, a decision was reached to use the 1990 data as scored by the 1990 raters.

Tables 15-4 and 15-5 show the results of the comparison of the rescore of the 1988 data with the scores assigned to the papers in 1988. Table 15-4 shows, by grade and item, the average difference between the 1990 rescore and the 1988 score and the standard deviation of the difference. Table 15-5 shows the distribution of the difference between the rescore and the original score, again by age and item. The average difference between the rescore and the original score is -0.031 for grade 4, 0.007 for grade 8 and -0.018 for grade 11.

In light of the slight differences between the 1988 and the 1990 scoring, and because the between variability for 1988 and 1990 and the within variability for 1990 scoring were quite similar, direct comparisons between the 1990 results and the original 1988 results were considered to be acceptable. Consequently, the 1988 trend point was based on 1988 scored data. The resultant sample sizes for the trend report analyses are given in Table 15-6. (The 1984 trend point was based on a rescoring of a sample of the 1984 data by the 1988 raters. For details, see Johnson, 1990.)

15.1.2 The Writing Trend Scale Based on the Average Response Model

Although analyses on individual items were also conducted, the initial plan was to use the average response model (ARM) of scaling nonbinary data (Beaton & Johnson, 1987, 1990) for the analysis of trends in writing achievement. The intention was to place the 1990 data onto the ARM writing trend scale established in 1988 and documented in Johnson (1990). However, because the 1990 results based on the ARM were inconsistent with the individual item level results, an alternative "meanparts" procedure was considered. After the results for the two procedures were compared, the meanparts procedure was chosen (for reasons explained in the following sections).

Table 15-3

Percentages of Exact Score Agreement and Interrater Reliability for the Primary Trait Scoring of the Writing Trend Items

NAEP Item	1988 Data (by 1988 Raters)		1990 Data (by 1990 Raters)		Rescore of 1988 Data (1990/1988 Raters)	
	Percent Agreement	Reliability	Percent Agreement	Reliability	Percent Agreement	Reliability
Grade 4/Age 9						
N000602 XYZ Company	97.1	.99	88.8	.83	91.1	.90
N000902 Radio Station	93.5	.95	92.1	.93	89.0	.90
N001002 Appleby House	90.3	.92	78.5	.72	76.9	.78
N007602 Flashlight	87.5	.88	78.2	.77	80.5	.74
N014702 Plants	94.3	.95	82.4	.86	88.5	.89
N014802 Spaceship	91.8	.95	75.2	.82	83.7	.89
Grade 8/Age 13						
N000302 Recreation Opportunity	85.4	.82	76.7	.73	83.0	.81
N000402 Food on Frontier	79.9	.68	72.1	.67	83.5	.78
N000502 Dissecting Frog	76.1	.64	66.1	.56	80.6	.70
N000602 XYZ Company	93.5	.92	86.8	.76	92.6	.87
N000902 Radio Station	87.0	.89	80.7	.83	82.0	.79
N001002 Appleby House	75.3	.69	75.9	.72	75.4	.75
Grade 11/Age 17						
N000302 Recreation Opportunity	90.8	.93	76.3	.78	71.6	.78
N000402 Food on Frontier	93.1	.86	76.7	.73	78.9	.69
N001002 Appleby House	89.3	.89	81.6	.82	81.1	.81
N018002 Spaceship	89.9	.93	71.8	.75	73.2	.75
N019002 Job Application	92.3	.92	84.6	.83	85.5	.86
N021002 Bike Lane	84.9	.87	75.6	.78	78.2	.76

Table 15-4
Mean and Standard Deviation of (Rescore - Original)
for the 25% Rescore of 1988 Writing Responses

NAEP Item	Grade 4		Grade 8		Grade 11	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
N000602 XYZ Company	-.078	.462	-.003	.408	—	—
N000902 Radio Station	-.012	.340	.022	.499	—	—
N001002 Appleby House	-.025	.480	.065	.501	-.025	.435
N007602 Flashlight	-.080	.487	—	—	—	—
N014702 Plants	.039	.376	—	—	—	—
N014802 Spaceship	-.040	.425	—	—	—	—
N000302 Recreation Opportunity	—	—	-.060	.444	-.015	.547
N000402 Food on Frontier	—	—	.022	.416	-.010	.468
N000502 Dissecting Frogs	—	—	.000	.467	—	—
N018002 Space Program	—	—	—	—	.003	.541
N019002 Job Application	—	—	—	—	-.027	.449
N021002 Bike Lane	—	—	—	—	-.032	.491
OVERALL	-.031		.007		-.018	

Table 15-5
Distribution of (Rescore-Original) for the 25% Rescore of 1988 Writing Responses
Percent of Responses Where (Rescore-Original) Equals -1, 0, or 1

NAEP Item	Grade 4			Grade 8			Grade 11		
	-1	0	1	-1	0	1	-1	0	1
N000602 XYZ Company	4.7	9[illegible]1	0.7	3.3	92.6	1.0	—	—	—
N000902 Radio Station	5.8	89.0	5.0	6.7	82.0	9.9	—	—	—
N001002 Appleby House	12.8	76.9	10.3	8.6	75.4	15.7	10.7	81.1	8.2
N007602 Flashlight	12.0	80.5	6.1	—	—	—	—	—	—
N014702 Plants	3.5	88.5	7.0	—	—	—	—	—	—
N014802 Spaceship	9.6	83.7	76.1	—	—	—	—	—	—
N000302 Recreation Opportunity	—	—	—	10.0	83.0	6.0	15.2	71.6	12.7
N000402 Food on Frontier	—	—	—	6.8	83.5	9.5	10.7	78.9	10.2
N000502 Dissecting Frogs	—	—	—	9.1	80.6	9.6	—	—	—
N018002 Space Program	—	—	—	—	—	—	13.1	73.2	12.9
N019002 Job Application	—	—	—	—	—	—	7.9	85.5	4.6
N021002 Bike Lane	—	—	—	—	—	—	11.3	78.2	9.7
OVERALL	7.9	85.2	5.8	7.4	82.9	8.6	11.5	77.9	9.8

Table 15-12

Sample Comparisons of Students Providing and Not Providing Portfolios, Age 9

Rating	Portfolio		Booklets 54-55, No Portfolio		Total	
	Weighted N	Col %	Weighted N	Col %	Weighted N	Col %
N000602 XYZ Company (Chi-square = 15.4192; P = 0.0087)						
0	67.0	6.29	86.8	9.68	153.8	7.84
1	281.1	26.36	236.0	26.33	517.0	26.34
2	21.1	1.98	24.3	2.71	45.4	2.31
3	166.8	15.64	100.2	11.18	267.0	13.60
4	0.0	0.00	0.0	0.00	0.0	0.00
Missing	530.5	49.74	449.1	50.10	979.5	49.91
Total	1066.4	100.00	896.2	100.00	1962.6	100.00
N000902 Radio Station (Chi-square = 34.7288; P = 0.0000)						
0	71.7	6.72	100.3	11.19	172.0	8.76
1	504.8	47.34	473.9	52.88	978.7	49.87
2	356.7	33.45	264.1	29.46	620.8	31.63
3	131.2	12.30	57.3	6.39	188.5	9.60
4	2.0	0.19	0.7	0.08	2.7	0.14
Missing	0.0	0.00	0.0	0.00	0.0	0.00
Total	1066.4	100.00	896.2	100.00	1962.6	100.00
N001002 Appleby House (Chi-square = 30.1934; P = 0.0000)						
0	261.2	24.50	287.6	32.09	548.9	27.97
1	192.8	18.08	183.3	20.45	376.1	19.16
2	512.5	48.06	383.9	42.83	896.4	45.67
3	97.9	9.18	41.5	4.62	139.4	7.10
4	1.9	0.17	0.0	0.00	1.9	0.09
Missing	0.0	0.00	0.0	0.00	0.0	0.00
Total	1066.4	100.00	896.2	100.00	1962.6	100.00
N014802 Spaceship (Chi-square = 24.3946; P = 0.0002)						
0	44.5	4.17	69.9	7.80	114.4	5.83
1	179.1	16.80	169.2	18.88	348.4	17.75
2	192.0	18.01	146.8	16.38	338.8	17.26
3	117.6	11.03	58.0	6.47	175.6	8.95
4	2.7	0.25	3.2	0.36	5.9	0.30
Missing	530.5	49.74	449.1	50.10	979.5	49.91
Total	1066.4	100.00	896.2	100.00	1962.6	100.00

Analyses were performed by grade to determine whether the sample of students from whom portfolios were obtained was representative of the assessed national writing sample. The results for each prompt, shown in Table 15-12 for age 9 and Table 15-13 for age 13, indicate that students who provided portfolios ("Portfolio") tended to demonstrate higher achievement in writing than students responding to booklets 54 and 55. Students who provided portfolios also tended to demonstrate higher achievement in writing than those who did not participate ("Booklets 54-55, No Portfolio").

Comparison analyses were made to determine differences between the two samples (portfolio and no portfolio) in terms of background variables. The results are presented in Table 15-14. For age 9/grade 4, the background variables on which the two samples tended to differ were modal age, modal grade, home environment, and television watching; for age 13/grade 8, the variables were size and type of community, modal grade, school type, homework, and grades in school.

The portfolio papers were scored using two sets of criteria. One set consisted of a descriptive set of criteria; the other consisted of an evaluative set of criteria. Both sets were derived from recommendations made by a committee of outside educators with expertise in the field and ETS staff.

Frequency distributions were performed on the results obtained from the descriptive and evaluative scoring. The scored portfolio data was linked to data from the trend writing assessment (primary trait, holistic, and mechanics). Relationships between the descriptive and evaluative scoring, as well as the writing assessment, were examined. A more detailed report is given in *The Writing Students Do in School: The 1990 NAEP Portfolio Study of Fourth and Eighth Graders' School-based Writing*.

Table 15-10

1990 Sample Sizes for Mechanics Scoring

Grade	NAEP Item	Sample Size
4	Spaceship	567
8	Recreation Opportunity	601
11	Recreation Opportunity	602

Table 15-11

Sample Sizes for Holistic Scoring

Grade	NAEP Item	Sample Size		
		1984*	1988*	1990
4	Flashlight	609	614	702
	Spaceship	611	1258	1367
8	Food on Frontier	603	1339	1503
	Recreation Opportunity	494	1372	1498
11	Food on Frontier	629	1212	1401
	Recreation Opportunity	521	1242	1415

* All 1984 and 1988 rescored papers were also holistically scored.

information more efficiently. However, because it is model-free the meanparts approach will produce results guaranteed to be consistent with the item-level results.

Because of the discrepancy between the item-level trend results and the ARM-based trend results, it was decided to use the model-free meanparts approach for the reporting of writing trend.

15.1.4 Other Analyses of Trends in Writing Performance

In addition to trends in primary trait scores, trends were also measured for the mechanics of writing and for overall writing fluency. Trends in components of the mechanics of writing at each age were based on a selected writing prompt given to the age group in 1984, 1988, and 1990. The writing items used for the assessment of the mechanics of writing were "Spaceship" (N014840) for grade 4 and "Recreation Opportunity" (N000310) for grades 8 and 11. All analyses were based on representative subsamples of around 500 responses to each item at each grade and year. In the sample selection, Black students were sampled at a higher rate to provide sufficient sample size to allow for comparisons in performance between Black and White students. The student weights were adjusted to reflect this oversampling of Black students by a poststratification process: For each grade, the students selected for the writing mechanics analysis were categorized by gender and by race/ethnicity (White, Black, Hispanic, other), producing eight cells. The sampling weights of the students within each cell were then multiplied by a poststratification factor computed as a ratio whose denominator is the sum of weights of all students in the cell selected for the mechanics analysis and whose numerator is the sum of the weights of all students in the writing assessment of the specified grade, gender, and race/ethnicity. All papers used in this analysis were scored in 1990; the actual sample sizes are shown in Table 15-10.

Two writing items for each grade in the bridge samples were holistically scored for overall writing fluency. To allow the measurement of trends in overall writing fluency, a sample of responses in the 1984 and 1988 assessment to the same items were also holistically scored. Table 15-11 shows the sample sizes for the measurement of trends in the fluency of writing.

15.2 WRITING PORTFOLIO ANALYSES

The writing portfolio study was conducted on a subset of the national writing sample; that is, only age 9/grade 4 and age 13/grade 8 students who were administered booklet 54 or 55 were asked to participate. At age 9/grade 4, the response rate was 54 percent; papers were received from 1,066 of the 1,962 students who were sampled for the school-based writing study. At age 13/grade 8, the response rate was 51 percent; papers were received from 1,059 of the 2,071 sampled students. The rather low participation rates were in part due to the method of collecting the portfolio data. Teachers were asked to mail in students' samples of writing, rather than handing them in to NAEP field administrators.

is an unbiased estimator of

$$S = \sum_{i=1}^{p} a_i S_i \quad .$$

It is also possible to obtain an estimate of the sampling variance of s^* by jackknifing the matrix $[s_1^*, ..., s_p^*]$ at the PSU level because, due to BIB spiraling, equivalent samples of the population of students within each PSU respond to each item. Let

$$\tau_k = [s_{1k}^*, ..., s_{pk}^*]$$

be the matrix with columns corresponding to the pseudoreplicates of the s_i^* corresponding to the k^{th} PSU pair. Then the pseudoreplicate of $s^* = [s_1^*,...,s_p^*]a$ corresponding to the k^{th} PSU pair is

$$s_k^* = \tau_k \, a$$

and the jackknife variance estimate of s^*, which accounts for interitem covariances is

$$\hat{Var}(s^*) = \sum_{k=1}^{M} (s_k^* - s^*)(s_k^* - s^*)'$$

which is a variance-covariance matrix of order r where r is the number of elements in the vector s^*.

Because the estimator s^* of group level data is computed as a linear combination of unbiased estimators of the corresponding parameters for each of the constituent part of $\Theta = Xa$, and because this linear combination is often a mean, the estimator s^* will be referred to as the *meanparts estimator*.

The meanparts estimator of some quantity of interest, say a group mean, differs from the equivalent estimator based on the ARM scale values in a fundamental way. The average response method seeks to obtain an unbiased estimate of the mean writing score for every individual (and goes further by also addressing the variability of that estimated score). If the method is successful, meaning that the model fits the data, then any statistics based linearly on these ARM plausible values are automatically unbiased.

On the other hand, the meanparts estimator never produces an estimate of an individual's scale value, but rather directly produces estimates of *aggregate* quantities, where it is required that those aggregates can be expressed as a linear combination of the equivalent aggregates of the constituent items. The advantage of this is that such estimates are unbiased and model-free. The disadvantage is that each separate analysis requires its own specific computation of the pertinent meanparts estimator, this computation requiring p separate computations: one for each of the items. This produces a considerable increase in the computational load required for exploratory analysis. Furthermore, the variance of the meanparts estimator can exceed that of the ARM estimator because the latter uses the available

Suppose that the value of the mean score across the p items were known for every individual in the sample, so that the vector Θ was completely known, and consider the statistic

$$s = L'\Theta,$$

for some vector or matrix L. Thus s is a linear combination of the elements in Θ. Examples of this are subgroup means, contrasts of subgroup means, and more generally, regression coefficients.

Suppose that s is an unbiased estimator of the population value S. Then, since

$$S = E(s) = E(L'\Theta) = E(L'X)a ,$$

the quantity of interest S can be expressed as a linear combination of component quantities, S_i, where S_i is the equivalent population value for the scores X_i on item i, and where S_i is estimated unbiasedly by the statistic

$$s_i = L'X_i .$$

(For the moment we assume that the score on item i is known for all individuals in the sample.)

As an example, if S is a vector of subgroup means of the average performance across the p items, then S_i is the vector of subgroup mean performance on the specific item i and so S is quite evidently the average of these item level mean performance vectors.

Now, although the score on item i is only known for a subsample of students, this subsample is a representative sample of the population. This means that an unbiased and consistent estimator of the item level parameter vector S_i based only on the available information from the subsample of students responding to the item is

$$s_i^* = L_i' X_i^*$$

where X_i^* is the vector of known scores and L_i' is the matrix of associated values, chosen so that

$$E(s_i^*) = S_i .$$

In the example where S_i is the r x 1 vector of r subgroup mean performance levels on item i, the corresponding estimator s_i^* is the r x 1 vector of the weighted mean scores, by subgroup, across all members of the subgroup responding to the item.

Then, since s_i^* is an unbiased estimator of S_i, for each item i, it follows automatically that

$$s^* = \sum_{i=1}^{p} a_i s_i^*$$

one item. (The partial correlation between N001002 and N018002 and N021002 went from 0.11 and 0.10 in 1988 to -0.02 and 0.05 in 1990—the partial correlations between N001002 and N019002 went from 0.05 to 0.06.) The result of the decline in the value of $\beta_{24 \cdot 14}$ is a lessened $\overline{X}_{2t}$ impact of the difference $\overline{X}_{1t}$ - $\overline{X}_{14}$ and a consequent overprediction of the value $\overline{X}_{2t}$.

Table 15-9
Components of the ARM Overall Proficiency by Grade:
Based on First Plausible Value

	Assessment Year	Grade 4 Mean Score	Grade 8 Mean Score	Grade 11 Mean Score
$\overline{X}$	1984	1.72	2.12	2.23
	1988	1.75	2.09	2.21
	1990	1.78	2.07	2.19
$\overline{X}$ In grade	1984	1.81	2.06	2.12
	1988	1.87	2.03	2.14
	1990	1.85	1.98	2.12
$\overline{X}$ Other lowest*	1984	1.52	2.45	2.52
	1988	1.52	2.36	2.44
	1990	1.50	2.35	2.49
$\overline{X}$ Other highest**	1984	1.76	2.03	2.25
	1988	1.79	2.02	2.19
	1990	1.95	2.06	2.14

* Other lowest for grade: 4 is grade 8, 8 is grade 4, 11 is grade 4
** Other highest for grade: 4 is grade 11, 8 is grade 11, 11 is grade 8

Because the ARM results were so strongly affected by predicted values based on a single item and because the trend results were inconsistent with the item level trend results, the decision was made to abandon the ARM-based estimates for reporting trend and to instead employ an alternative, model-free procedure.

15.1.3 Meanparts Summarization: An Unbiased Estimator for Combinations of Mean Scores

An alternative estimation of a combination of scores is based on the facts that:

1) the target quantity of interest, Θ, is the vector of θ_i, the values of the composite for each student. Θ is a linear combination of the vectors $X_1, X_2, ..., X_p$, where X_i is the vector of scores on item i. In particular, $\Theta = Xa$, where $X = [X_1, X_2, ..., X_p]$, and

2) the information on the values of each of the item score variables, the X_i, is available on a representative subsample of the population.

is unbiased under the assumption that $[X_{1d}\ X_{2d}]$ and $[X_{1t}\ X_{2t}]$ are each normally distributed with common variance matrix Σ and mean matrices $Y_d B$ and $Y_t B$, where B is a matrix.

In particular, the predicted value of the target population mean performance on the set of items X_2 is

$$\hat{\bar{X}}_{2t} = \bar{X}_{2d} + (\bar{X}_{1t} - \bar{X}_{1d})\beta_{2d\cdot 1d} + (\bar{y}_t - \bar{y}_d)\delta^{-1} y_{d\cdot 1}(\bar{X}_{2d\cdot 1} - X_{1d\cdot 1}\beta_{2d\cdot 1d}) \tag{15.6}$$

where $\bar{X}_{2d}$ is the mean performance on the set of items X_2 for the donor population; $\bar{X}_{1t}$ and $\bar{X}_{1d}$ are the mean performance on the X_1 items for the target and donor populations; $\bar{y}_t$ and $\bar{y}_d$ are the means of the columns of Y_t and Y_d excluding the intercept column (so that $Y_t = [\underline{1}y_t]$ and $Y_d = [\underline{1}y_d]$);

$y_{d\cdot 1} = y_d - \underline{1}\,\bar{y}_d$

$X_{2d\cdot 1} = X_{2d} - \underline{1}\,\bar{X}_{2d}$ and

$X_{1d\cdot 1} = X_{1d} - \underline{1}\,\bar{X}_{1d}$

are the columns of y_d, X_{2d}, and X_{1d} centered by their means;

$\delta_d = y_{d\cdot 1}'y_{d\cdot 1}$; and

$\beta_{2d\cdot 1d} = (X_{1d\cdot Y}'X_{1d\cdot Y})^{-1}X_{1d\cdot Y}'X_{2d\cdot Y}$ where

$X_{1d\cdot Y} = (I - Y_d(Y_d')^{-1}Y_d'Y_d)X_{1d}$

is X_{1d} linearly adjusted for Y_d and $X_{2d\cdot Y}$ is similarly defined. $\beta_{2d\cdot 1d}$ is the coefficient vector of the regression of X_{2d} on X_{1d} after linearly adjusting for the conditioning variables Y_d.

Table 15-9 shows the overall average ARM proficiency estimate by grade and year as well as three components of that estimate. (All estimates in this table are based on a single plausible value and so are subject to some variability due to imprecision of individual measurement.) The first component, labeled "$\bar{X}$ In grade," gives the means based on the items presented to that grade—these means correspond exactly to the observed averages across the items presented to the grade and shown in Table 15-8. The next component, labeled "$\bar{X}$ Other lowest," shows the predicted means for the items presented only to grade 8 for the grade 4 column and to grade 4 for the grade 8 and grade 11 columns. The final component, "$\bar{X}$ Other highest," shows the predicted means for the items presented only to grade 11 for the grade 4 and grade 8 columns and to grade 8 for the grade 11 column. The table shows that the apparent increase in ARM-based writing proficiency at grade 4 is due solely to the substantial increase in the predicted performance on the grade 11 items. This is largely due to a decline in the value of $\beta_{2d\cdot 1d}$ of (15.5) in 1990 relative to 1988. The prediction of grade four performance on the three unique grade 11 items (N018002, N019002, and N021002) is based on their linkage with a single item (N001002) that appears in both grade 4 and grade 11. The partial correlation between the linking item and the grade 11 unique items (partialing out conditioning variables) declined substantially between 1988 and 1990 for two items and stayed essentially the same for

The procedures for estimating each of these three types of elements are given in Johnson (1990).

It was the estimation of Type 2 elements that lead to the anomalous results for the 1990 ARM-based trend estimates. Of the 11 writing tasks that made up the ARM composite, two were presented only at grade 4, one only at grade 8, and three only at grade 11. Two items were presented at both grades 4 and 8, two items at both grades 8 and 11 and one item was presented at all three grades. The prediction of performance for a given grade (e.g., grade 4) on items given at another grade (e.g., grade 8) is accomplished using regression based techniques where it is assumed that the conditional distribution on the exercises presented only at the other grade (grade 8), given the conditioning variables and the performance on the exercises presented both to the given grade (grade 4) and the other grade (grade 8), is the same for both grades.

As described above, a Type 2 element of the cross-product matrix for a given grade corresponds to a term that is not directly estimable based on data for that grade (the *target* grade) but that can be estimated based on data from another grade (the *donor* grade). The estimation procedure is as follows.

Let X_1 be the set of items held in common between the target grade and the donor grade and let X_2 be the set of items presented to the donor but not to the target. For notational convenience, we will operate as if the entire donor and target population had been measured and that complete information by student is available for all items presented to the student's grade and year. There is no loss of generality because only estimates of the terms of the cross-product matrix are required.

The known information for the target population is the matrix

$$V_t = \left[Y_t \;\; X_{1t}\right]$$

consisting of the conditioning variables and the items held in common with the donor population. The known information for the donor population is the corresponding matrix

$$V_d = \left[Y_d \;\; X_{1d}\right]$$

plus the set of items X_{2d}.

We seek estimates of the Type 2 terms $V_t'X_{2t}$ and $X_{2t}'X_{2t}$ in the cross-product matrix

$$C_t = \begin{bmatrix} T_t'V_t & V_t'X_{2t} \\ X_{2t}'V_t & X_{2t}'X_{2t} \end{bmatrix} .$$

Beaton and Johnson (1987) show that the estimator

$$\hat{V_t'X_{2t}} = (V_t'V_t)\,(V_d'V_d) - 1V_d'X_{2d}$$

ϵ_{ik} is a random draw from a $N(O,\sigma_\epsilon^2)$ distribution, where σ_ϵ^2 is the residual mean-squared-error for the regression defined by (15.4). The vector

$$\begin{bmatrix} \gamma_k \\ \alpha_k \end{bmatrix}$$

is a draw from a multivariate normal distribution with mean O and variance-covariance matrix

$$\Sigma = (V_1'V_1)^{-1}\,\sigma_\epsilon^2.$$

The values of α_k and γ_k are held fixed for all students with the same pattern of missing data.

A further discussion of the generation of ARM plausible values, given an estimate C of the sum of squares and cross product matrix Y'V, appears in Beaton and Johnson (1990). The next subsection considers the estimation of V'V.

15.1.2.2 Estimation of V'V

As noted in the previous subsection, the basis for the estimation of a predicted value for any student is an estimate C of the full sums-of-squares-and-cross products matrix

$$V'V = \begin{bmatrix} Y'Y & Y'X \\ X'Y & X'X \end{bmatrix} \qquad (15.5)$$

from which all other necessary matrices and estimates are derived. For the construction of the NAEP writing trend scale, nine separate estimates of the cross-product matrix were created: one for each of the three grades for each of the years 1984, 1988, and 1990. The elements of the estimate C of V'V for a particular grade and year fall into three general types:

Type 1: Elements that are directly estimable from the available data for that grade and year; these are sums of squares and cross-products involving the conditioning variables and the items presented to that grade in that year.

Type 2: Elements that must be estimated based on relationships observed for another grade in the same year or for the same grade in the other year; these are sums of squares and cross-products involving items and pairs of items not administered to the target grade and year but administered to another grade or year.

Type 3: Elements requiring the imputation of between-item correlations; these are cross-product terms involving pairs of items that have never been presented together so that the between-item correlation is not estimable.

where the elements of T, θ_i, are the values of the composite for each student in the population. The exact value of θ_i will not be known unless the student i was administered all p of the exercises. The plausible values, $\tilde{\theta}_{ik}$, of equation (15.1) are determined by operations on the matrix C_θ where C_θ is the estimated population sum of squares and cross product matrix of the conditioning variables, the writing exercises and the composite. C_θ is generated by the matrix C and the transformation matrix

$$H = \begin{bmatrix} I_q & O & O \\ O & I_p & a \end{bmatrix}$$

by

$$C_\theta = H'CH = \begin{bmatrix} Y'Y & Y'X & Y'T \\ X'Y & X'X & X'T \\ T'Y & T'X & T'T \end{bmatrix} \tag{15.3}$$

The matrix C_θ can be used to estimate a value of θ_i for student i as follows: Let X_1 consist of the columns of X corresponding to the writing exercises presented to student i and let V_1 = [Y X_1]. The least-squares estimates of $\hat{\beta}$ and $\hat{\Gamma}$ in (15.1) are

$$\begin{bmatrix} \hat{\Gamma} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} Y'Y & Y'X_1 \\ X_1'Y & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} Y'T \\ X_1'T \end{bmatrix} \tag{15.4}$$

and the standard least-squares point estimate of the composite score for student i is

$$\hat{\theta}_i = x_i\hat{\beta} + y_i\hat{\Gamma} .$$

This value is the mean of the predictive distribution of potential θs for the individual and, thus, does not take into account the fact that any other value from this predictive distribution might also have been the student's score. By including the terms accounting for the uncertainty in the estimation of a student's composite score, the plausible values $\tilde{\theta}_{ik}$ allow the more complete representation of what is known and what is not known about the student's composite scores. The terms accounting for uncertainty are of two types:

1) ϵ_{ik}, accounting for variability of potential scores of an individual about the conditional mean (of the distribution given y_i and x_i) and

2) α_k and γ_k, accounting for uncertainty due to using sample estimates of $\hat{\beta}$ and $\hat{\Gamma}$ in the regression equation.

Let x_i represent the (row) vector of responses of the ith student to the questions in the ARM composite that were presented to that student and let y_i be the (row) vector of values of that student's conditioning variables. Then a plausible value from the conditional distribution of θ given the observed data x_i and y_i for student i is

$$\tilde{\theta}_{ik} = x_i\hat{\beta} + y_i\hat{\Gamma} + x_i\alpha_k + y_i\gamma_k + \epsilon_{ik} \tag{15.1}$$

where

$\tilde{\theta}_{ik}$ is the k[th] plausible value of the ARM composite

$\hat{\beta}$ is the (column) vector giving the change in the composite for unit change in the scores on each of the questions in x_i

$\hat{\Gamma}$ is the (column) vector of effects for the conditioning variables

α_k and γ_k are random draws from a multivariate normal distribution with mean vector 0 and variance-covariance matrix Σ where Σ is the variance-covariance matrix of the parameter estimates $\hat{\beta}$ and $\hat{\Gamma}$. (α_k and γ_k reflect the uncertainty due to using sample estimates $\hat{\beta}$ and $\hat{\Gamma}$ in the regression equation.)

ϵ_{ik} is an estimated residual drawn from a normal distribution with mean 0 and variance σ_ϵ^2 where σ_ϵ^2 is the variance of the predictive distribution of θ given the observed values of x_i and y_i.

All parameters in equation (15.1) were estimated by least-squares technology. To accomplish this, it is sufficient to obtain estimates of the elements of the population sum of squares and cross products matrix of the conditioning variables and the writing questions:

$$C = \textit{an estimate of } V'V = \begin{bmatrix} Y'Y & Y'X \\ X'Y & X'X \end{bmatrix} \tag{15.2}$$

In the above, Y is a $N \times q$ matrix containing the values of the q conditioning variables for each of the N students in the population; X is a $N \times p$ matrix containing the scores of the N students in the population on the p exercises; and $V = [Y\ X]$. If Y and X were known for all students in the population, C would be trivially equal to V'V. However, since only a sample of the students in the population were assessed for writing and since each sampled student was only presented a few writing questions, many of the elements of Y and X are unknown. Accordingly, V'V must be estimated. A description of the procedures used to determine an estimate C of V'V is given in section 15.1.2.2.

Since the ARM composite is the mean of the individual questions, the estimate C generates a complete set of sufficient statistics (the normal equations) for the standard least-squares prediction of an ARM composite value given conditioning variable characteristics and responses to any subset of writing questions. Define the N element column vector T by

$$T = Xa$$

Table 15-8
Writing Mean Scores for Items by Grade over Assessment Years

NAEP Item	Year	Mean Score and SEM		
		Grade 4	Grade 8	Grade 11
N000302 Recreation Opportunity	1984 1988 1990	— — —	1.58 (0.04) 1.49 (0.02) 1.48 (0.03)	1.90 (0.04) 1.86 (0.03) 1.86 (0.03)
N000402 Food on Frontier	1984 1988 1990	— — —	1.88 (0.03) 1.91 (0.03) 1.84 (0.02)	1.97 (0.03) 2.03 (0.02) 2.00 (0.02)
N000502 Dissecting Frogs	1984 1988 1990	— — —	1.94 (0.02) 1.98 (0.02) 1.86 (0.02)	— — —
N000602 XYZ Company	1984 1988 1990	1.80 (0.05) 1.77 (0.04) 1.80 (0.03)	2.57 (0.03) 2.49 (0.03) 2.47 (0.03)	— — —
N000902 Radio Station	1984 1988 1990	1.48 (0.03) 1.59 (0.03) 1.59 (0.03)	2.04 (0.03) 1.91 (0.02) 1.90 (0.03)	— — —
N001002 Appleby House	1984 1988 1990	1.79 (0.04) 1.95 (0.02) 1.86 (0.02)	2.37 (0.03) 2.39 (0.02) 2.32 (0.02)	2.41 (0.03) 2.44 (0.03) 2.42 (0.02)
N007602 Flashlight	1984 1988 1990	1.71 (0.03) 1.78 (0.04) 1.76 (0.04)	— — —	— — —
N014702 Plants	1984 1988 1990	2.22 (0.03) 2.22 (0.03) 2.12 (0.03)	— — —	— — —
N014802 Spaceship	1984 1988 1990	1.77 (0.04) 1.82 (0.02) 1.88 (0.03)	— — —	— — —
N018002 Space Program	1984 1988 1990	— — —	— — —	2.00 (0.04) 2.05 (0.03) 2.06 (0.02)
N019002 Job Application	1984 1988 1990	— — —	— — —	2.54 (0.03) 2.57 (0.03) 2.52 (0.02)
N021002 Bike Lane	1984 1988 1990	— — —	— — —	1.91 (0.03) 1.90 (0.03) 1.84 (0.02)

Table 15-7

Averages of Writing Mean Scores for Items by Grade over Assessment Years

NAEP Item	Year	Mean Score and SEM		
		Grade 4	Grade 8	Grade 11
Average--All items (excluding N007602)*	1984 1988 1990	1.81 (0.02) 1.87 (0.02) 1.85 (0.02)	2.06 (0.01) 2.03 (0.01) 1.98 (0.01)	2.12 (0.02) 2.14 (0.01) 2.12 (0.01)
Average--Common items for grades 4 and 8	1984 1988 1990	1.68 1.77 1.75	2.33 2.26 2.23	— — —
Average--Common items for grades 8 and 11	1984 1988 1990	— — —	1.94 1.93 1.88	2.09 2.11 2.09
Average--Common items for all grades	1984 1988 1990	1.79 (0.04) 1.95 (0.02) 1.86 (0.02)	2.37 (0.03) 2.39 (0.02) 2.32 (0.02)	2.41 (0.03) 2.44 (0.03) 2.42 (0.02)
ARM/100 results	1984 1988 1990	1.70 (0.02) 1.73 (0.01) 1.77 (0.01)	2.12 (0.01) 2.08 (0.01) 2.05 (0.01)	2.23 (0.02) 2.21 (0.01) 2.18 (0.01)

* Item N007602 was excluded from the ARM model since it never appeared with any other item.

Table 15-7 shows the mean score results for various groupings of writing items as well as the average ARM proficiency estimates (divided by 100) by grade for each of the years 1984, 1988, and 1990, along with the standard errors of these estimates. Table 15-8 shows the mean score results for individual writing items. The ARM results suggest that the fourth graders performed slightly higher in the 1990 assessment than in the 1988 assessment. On the other hand, based on the mean of the scores given only to fourth-grade students (also shown in the table), the performance based on 1990 data was slightly lower than the performance based on 1988 data. The discrepancy appears to be due to the prediction of performance on other than fourth-grade items, in particular, the predicted performance on items presented only to grade 11. (This will be discussed in more detail below.) Consequently, it was decided to apply an alternative procedure called *meanparts summarization*, the advantage being that it avoids making inferences with regard to students' writing achievement at a given grade on items not administered at that grade. In effect, meanparts summarization provides an estimate of average writing achievement across all items presented to a given grade. Further details are given in section 15.1.3. The rest of this section describes the ARM, which was used for previous assessments and planned to be used for the 1990 assessment.

15.1.2.1 Overview of the Average Response Method

The average response method begins with a defined composite of the (primary trait) scores for a set of p exercises and provides, for each assessed student, draws from the distribution of potential values for that composite. If a student had responded to all the exercises going into the composite, then that student's ARM score would be directly calculable, without error, by

$$\theta = a'x$$

where x is the vector of the subject's scores on the p questions in the composite and a is a vector of p arbitrary constants. For the ARM writing scale, the arbitrary constants are each equal to $1/p$, since the ARM writing scale is defined as the predicted average performance across the set of p writing questions.

Because each respondent is presented only a subset of the questions, the respondent's composite value is unknown and so must be estimated. Such an estimate is provided by the ARM technology. Briefly, the ARM technology is a kind of multiple regression that produces for each student a set of plausible values, each of which predicts what that student's composite score might plausibly be, based on the student's scores on the exercises in the composite that were presented to the student and based on the student's status on a selected set of background variables, called the conditioning variables.

The writing tasks to be included in the ARM scale comprised 11 of the 12 tasks shown in Table 15-2. One task, "Flashlight" (N007602), was excluded from the ARM scale since that item never appeared with any other writing item in any of the assessment years 1984, 1988, or 1990. Consequently, the correlation between responses to that item and any other item were inestimable.

Table 15-6

Sample Sizes for Primary Trait Analyses of Trends in Writing Performance

NAEP Item	Grade 4			Grade 8			Grade 11		
	1990 Bridge	1988 Bridge	1984 Rescore	1990 Bridge	1988 Bridge	1984 Rescore	1990 Bridge	1988 Bridge	1984 Rescore
N000602 XYZ Company	1288	1152	544	1487	1334	616	—	—	—
N000902 Radio Station	1386	1234	585	1512	1364	612	406	—	—
N001002 Appleby House	1111	925	530	1396	1256	588	1277	1041	599
N007602 Flashlight	697	611	609	—	—	—	—	—	—
N014702 Plants	1416	1285	656	—	—	—	—	—	—
N014802 Spaceship	1367	1258	611	—	—	—	—	—	—
N000302 Recreation Opportunity	—	—	—	1498	1372	494	1415	1242	521
N000402 Food on Frontier	—	—	—	1503	1339	603	1401	1212	629
N000502 Dissecting Frogs	—	—	—	1518	1356	641	—	—	—
N018002 Space Program	—	—	—	—	—	—	1451	1195	632
N019002 Job Application	—	—	—	—	—	—	1169	1424	603
N021002 Bike Lane	—	—	—	—	—	—	1424	1178	636

Table 15-13

Sample Comparisons of Students Providing and Not Providing Portfolios, Age 13

Rating	Portfolio		Booklets 54-55, No Portfolio		Total	
	Weighted N	Col %	Weighted N	Col %	Weighted N	Col %
N000602 XYZ Company (Chi-square = 16.4365; P = 0.0057)						
0	12.4	1.17	30.4	3.01	42.8	2.07
1	108.8	10.27	129.5	12.79	238.3	11.50
2	35.2	3.33	44.9	4.44	80.2	3.87
3	367.6	34.70	304.5	30.09	672.1	32.45
4	0.0	0.00	0.0	0.00	0.0	0.00
Missing	535.4	50.54	502.7	49.67	1038.1	50.11
Total	1059.4	100.00	1012.3	100.00	2071.4	100.00
N000902 Radio Station (Chi-square = 9.6950; P = 0.0843)						
0	4.5	0.42	9.5	0.94	14.0	0.67
1	177.5	16.76	197.8	19.54	375.3	18.12
2	199.4	18.82	192.4	19.01	391.7	18.91
3	131.6	12.42	106.3	10.50	237.8	11.48
4	11.1	1.04	3.4	0.34	14.5	0.70
Missing	535.4	50.54	502.7	49.67	1038.1	50.11
Total	1059.4	100.00	1012.0	100.00	2071.4	100.00
N001002 Appleby House (Chi-square = 3.6176; P = 0.6058)						
0	48.2	4.55	59.6	5.89	107.9	5.21
1	49.2	4.64	48.8	4.82	97.9	4.73
2	247.3	23.34	249.7	24.67	497.0	23.99
3	169.7	16.01	143.9	14.22	313.6	15.14
4	9.7	0.92	7.3	0.72	17.0	0.82
Missing	535.4	50.54	502.7	49.67	1038.1	50.11
Total	1059.4	100.00	1012.0	100.00	2071.4	100.00
N000302 Recreation Opportunity (Chi-square = 13.3384; P = 0.0204)						
0	28.3	2.67	51.4	5.07	79.6	3.84
1	333.5	31.48	304.7	30.10	638.2	30.81
2	138.7	13.09	119.1	11.77	257.8	12.44
3	30.3	2.86	27.6	2.73	57.9	2.80
4	4.6	0.43	0.0	0.00	4.6	0.22
Missing	524.0	49.46	509.3	50.33	1033.4	49.89
Total	1059.4	100.00	1012.0	100.00	2071.4	100.00

Table 15-13 (continued)

Sample Comparisons of Students Providing and Not Providing Portfolios, Age 13

Rating	Portfolio		Booklets 54-55, No Portfolio		Total	
	Weighted N	Col %	Weighted N	Col %	Weighted N	Col %
N000402 Food on Frontier (Chi-square = 7.8032; P = 0.1675)						
0	29.3	2.76	44.0	4.35	73.2	3.54
1	159.4	15.05	161.8	15.99	321.2	15.51
2	270.1	25.50	244.2	24.13	514.3	24.83
3	76.6	7.23	52.8	5.21	129.3	6.24
4	0.00	0.00	0.0	0.00	0.0	0.00
Missing	524.0	49.46	509.3	50.33	1033.4	49.89
Total	1059.4	100.00	1012.0	100.00	2071.4	100.00
N000502 Dissecting Frogs (Chi-square = 24.0509; P = 0.0002)						
0	6.8	0.64	24.8	2.45	31.5	1.52
1	129.4	12.21	165.2	16.32	294.5	14.22
2	333.6	31.49	285.2	28.18	618.8	29.87
3	52.5	4.96	31.3	3.09	83.8	4.05
4	1.8	0.17	2.9	0.28	4.7	0.22
Missing	535.4	50.54	502.7	49.67	1038.1	50.11
Total	1059.4	100.00	1012.0	100.00	2071.4	100.00

Table 15-14.

Comparison Analysis of Students Responding to Booklets 54 and 55 and Providing a Portfolio with Those Who Did Not

Variables	Age 9			Age 13		
	Chi-square	DF	P-value	Chi-square	DF	P-value
Gender	0.219	2	0.897	0.006	2	0.997
Race/Ethnicity	9.322	6	0.156	5.283	6	0.508
Parent Education	9.267	5	0.099	6.910	5	0.227
Size and Type of Community	15.321	7	0.032	56.863	7	0.000
Modal Age	84.594	3	0.000	2.829	3	0.419
Modal Grade	503.036	3	0.000	76.869	3	0.000
School Type	2.082	5	0.838	33.396	5	0.000
Home Environment	17.837	3	0.000	8.117	3	0.044
Television Watching	16.223	7	0.023	9.553	7	0.215
Homework	10.608	5	0.060	21.249	5	0.001
Grades in School	16.101	9	0.065	26.537	9	0.002
Pages Read	9.974	5	0.076	2.276	5	0.810
Number of Reports Written	15.953	8	0.043	12.535	8	0.129

PART III

Statistical Summary of 1990 NAEP Data

Chapter 16

STATISTICAL SUMMARY OF THE 1990 NAEP SAMPLES AND ESTIMATES OF THE PROFICIENCIES OF AMERICAN STUDENTS[1]

Bruce A. Kaplan

Educational Testing Service

The analysis of the 1990 NAEP data has resulted in the production of many thousands of tables presenting estimates of the proficiency of students, and various subgroups of students, in American schools. This chapter gives some selected results from the assessment as well as a statistical summary of the 1990 NAEP samples. The chapter assumes a general familiarity with the structure of NAEP as summarized in the Introduction and the overview Chapters 1 and 9.

Three of the many types of NAEP results are presented here:

- results of the instrument development process, including the sizes of the item pools and numbers of booklets;
- results of the sampling process, including the numbers of students in each sample by selected subgroups; and
- results of the parameter estimation process, including estimates of the proficiencies of several populations of students in reading, mathematics, science and writing.

Interpretive results from the estimates presented here have been reported in the NAEP subject area trend and cross-sectional reports. The 1990 secondary-use data files and user guide (Rogers, Kline, Johnson, Mislevy, Allen, & Rust, 1992) are available for those who wish to estimate other parameters of student performance from the NAEP data or to search for possible explanations for the population characteristics that are reported here.

The technical details of the estimation process that underlie these tables are covered in previous parts of this report and not repeated here. A detailed discussion of how to read and use the tables of background and proficiency results is given by Zwick (1987b).

[1] Information for various tables in this chapter was provided by Yim Fai Fong, David Freund, Steven Isham, Laura Jerry, Edward Kulick, Michael Narcowich, and Keith Rust.

16.1 MEASUREMENT INSTRUMENTS

For the 1990 assessment, 36 different assessment booklets and questionnaires were printed for age class 9, 33 for age class 13, and 38 for age class 17. These instruments are shown by age level and type in Table 16-1.

The item pool used to develop all bridge and cross-sectional booklets is described in Table 16-2. In general, there are two types of items, cognitive and noncognitive. The cognitive items are developed to measure proficiency in particular subject areas, such as reading and mathematics. Cognitive items may be open-ended or multiple-choice. The noncognitive items are usually questions about the student's or teacher's backgrounds and attitudes but may also probe other areas such as school policies or teaching methods. Because many items were used at more than one age class, the total number of items in an item pool is not the sum of the item pools used for the three age classes.

Table 16-3 shows the number of cognitive items in each subject area that were used in the separate samples.

The excluded student, teacher, and school questionnaires contained only noncognitive questions. The number of items in the noncognitive pools is the same as the number of items in the questionnaires. More information about the instruments that were developed is provided in Chapters 2 and 4.

16.2 SAMPLE CHARACTERISTICS

In this section, the characteristics of the final NAEP sample is described. The process by which the sample was selected is discussed in Chapter 3.

In the 1990 main assessment, NAEP contacted 1,696 schools, of which 1,237 contributed data to the assessment. The disposition of these schools is shown in Table 16-4. Some of the schools were unwilling to cooperate; others were believed to be eligible from the sampling frame, but were not. The cooperation rate is calculated as the sum of cooperating schools and the schools that were found to have no eligible students divided by the same sum plus the schools that refused or were from districts that refused to cooperate.

Table 16-4 also shows the number of schools in several categories: region of the country (Northeast, Southeast, Central, West), school governance (public, private, Catholic, Bureau of Indian Affairs, Department of Defense), size and type of community, degree of urbanicity, grade span of school, number of teachers, and number of students.

For the 1990 bridge (trend) studies, NAEP contacted 1,124 schools, of which 862 contributed data to the various bridge assessments. Table 16-5 supplies the same information for the schools assessed for the bridge studies that the previous table supplies for the main assessment schools.

The numbers of respondents to the teacher questionnaire are summarized in Table 16-6. The first column in this table includes the number of teachers who responded by age class and subject area. The second column is the number of students who were not linked to teachers. The

third column is the number of students linked to teachers, but not specific classes of these teachers. The last column is the number of students linked to their teachers and their specific classes. (See section 9.2.5 in Chapter 9 for a discussion of teacher/student matching.)

NAEP is administered in units called assessment sessions. If the number of students attending an assessment session is less than a predetermined number, the students missing from the session are assigned to a makeup session and then assessed. Table 16-7 shows the number of regular and makeup sessions in 1990 NAEP by age class for the main NAEP and two bridge samples. Altogether, 155,655 students were involved in the 1990 NAEP, including excluded students. The breakdown by age class and by sample is shown in Table 16-8.

Tables 16-9 through 16-11 display the distribution of the students assessed in the main NAEP assessment in several basic categories for the three age classes: gender, racial/ethnic grouping, region of the country, parental education, and size and type of community. These tables have four columns:

- eligible by age, which means that the students were in an appropriate age group;
- eligible by grade, which means that the students were in an appropriate grade;
- eligible by age and by grade, which means that the students were of both an appropriate age and appropriate grade; and
- eligible by age or by grade, which is the total number of students for whom data were collected.

Tables 16-12 through 16-21 contain the distribution of students in the same categories by age class for the bridge samples. Tables 16-12 to 16-14 contain the distributions for the bridge to 1984 sample. Tables 16-15 to 16-17 display the distributions for the bridge to 1986 age only sample. Table 16-18 displays the distributions for the bridge to 1986 for the age 17/grade 11 sample. Table 16-19 and Table 16-21 enumerate the students in the three age classes assessed as part of the long-term bridge.

Similarly, Tables 16-22 through 16-24 contain the distribution of excluded students by age class for the main sample. Tables 16-25 through 16-27 enumerate the excluded students across the various bridge samples.

16.3 POPULATION ESTIMATES

The 1990 NAEP samples were designed for estimating the size and attributes of a number of different populations of students. The estimation procedures use sampling weights, developed by Westat, Inc., that are used in conjunction with the members of the sample (see Chapter 3). In this chapter, all estimates of population parameters use these sampling weights.

Table 16-28 shows the sizes of the various samples and the estimated population sizes by age/grade. (The sum of the initial weights for a given sample is an estimate of the number of students who are in the population represented by the sample. In other words, the sum of the

initial weights is taken as the estimated population size. In analyses, however, this sum of weights were rescaled to sum to the sample size).

Due to design considerations the main assessment was divided into subsamples, and were administered, and therefore weighted, independently, so that the sum of the initial weights for each subsample estimates the population size. Subsamples were formed by season and by subject area as follows: Reading and science formed one group, mathematics formed a second, and special mathematics formed a third. This yielded six subsamples (three groupings for each of two seasons, winter and spring). In calculating the estimate of population size across subsamples, a factor proportional to the number of booklets was used. The proportions for the reading and science subsamples for both spring and winter were 17/50 for age 9 and 14/44 for ages 13 and 17. The proportions for mathematics subsamples, regardless of season, were 7/50 for age 9 and 7/44 for ages 13 and 17. The proportions for special mathematics subsamples, regardless of season, were 1/50 for age 9 and 1/44 for ages 13 and 17.

Table 16-28 also lists winter and spring samples separately. The factors used for accumulation were twice those used for the combined sample estimates of the population size.

Note that the samples for the main assessment, the samples for all three age classes of the bridge to 1984, and the samples for the bridge-to-1986 age/grade sample are grade and age samples. The samples for the bridge-to-1986 age-only samples and long-term trend samples are age-only samples. The sum of the initial weights of the excluded students estimates the number of ineligible students at the respective age/grade levels.

In most cases, the number of students in an age/grade combination is not of interest; a researcher will be interested in estimating the number of students at either a grade or an age level. For the samples that contain both grade- and age-eligible students, an estimate of the total number of students at an age level can be made by summing the initial weights of only the age-eligible students and adding the corresponding sample of age-eligible excluded students' initial weights. An estimate of the total number of students in a grade sample can be made by summing the initial weights of grade-eligible students plus the initial weights of grade-eligible students from the appropriate excluded student sample.

The next group of tables estimates how many students in the main NAEP samples are age-eligible and grade-eligible by age class. Tables 16-29 through 16-31 show how many students at a particular grade level are at, in, or above the modal age for that grade, and how many at a particular age level are at, in, or above the modal grade for that age. Along with the counts from these samples are estimates of the numbers of students in these categories in the population. The standard errors of these estimates and coefficients of variation are also given. (The coefficient of variation of the estimated population size is defined as 100 times its standard error divided by the estimated population size.)

Tables 16-32 through 16-37 contain similar information for the bridge booklets, by age level. Where age-only samples are shown, information for all three ages is given in one table, since the partitioning of the sample by modal age groupings provides no added information.

Tables 16-38 to 16-56 show the sizes of the estimated populations of assessable students and the weighted percentages for the NAEP reporting categories of gender, race/ethnicity, region of the

country, parents' education level, and size and type of community. The estimated subpopulation percentages for the main NAEP samples are shown in Tables 16-38 through 16-40, separately by age eligibility, grade eligibility, and age/grade eligibility. Tables 16-41 to 16-50 show the same information for the bridge samples. In a similar manner, Tables 16-51 to 16-56 show the estimated total population of excluded students and the weighted percentages by demographic subgroups (data about parents' education level is not collected for excluded students and therefore not reported).

Students were assigned proficiency values in a subject area only if they received at least one assessment block in that area. Thus, the sample sizes of students who have proficiency values vary from one subject area to another. Tables 16-57 through Table 16-61 show the number of students with proficiency values in each subject area by age and grade combinations.

The rest of the tables in this chapter provide selected proficiency results for students sampled in the 1990 assessment. Tables 16-62 to 16-79 contain population estimates of student proficiencies by grade and by the subpopulations of gender, race/ethnicity, and parents' education level. The information about proficiency includes the mean and standard deviation of each subpopulation as well as the value of the 5th, 10th, 25th, 50th (median), 75th, 90th, and 95th percentiles. Results are shown separately for each subject area. Standard errors of the estimates are included in parentheses.

Tables 16-80 through 16-115 contain results for more finely defined subpopulations. Three of the major reporting variables—gender, race/ethnicity, and parents' education level—are cross-classified with one another (for example, Table 16-81 cross-classifies gender, race/ethnicity, and parents' education level with the race/ethnicity grouping for fourth-graders in the main reading sample). The data from these and other cross-classifications were used in the creation of the 1990 subject-area reports. Information provided for subpopulations includes the actual sample size (N); the estimated proportion of the population and its standard error (WEIGHTED PCT); the coefficient of variation of the estimated population size (<CV>); the proportion of students in each subpopulation and its standard error; and the average proficiency of the students and its standard error (shown directly below the corresponding proportion).

Table 16-1

Measurement Instruments Developed for 1990 NAEP

	Age Class		
	9	13	17
Student Assessment Booklets			
Main and Special	25	22	22
Bridge to 1984	6	6	6
Bridge to 1986--age only	3	3	2
Bridge to 1986--grade/age	0	0	6
Long-term Trend	2	2	2
Total Unique Booklets	**36**	**33**	**38**
Questionnaires			
Excluded Student	1	1	1
Teacher	1	2	0
School	1	1	1

Table 16-2

Number of Items Administered, by Age Class

	Age Class			Total
	9	13	17	Distinct Items
Common Background	66	71	107	119
Reading				
Background	133	144	202	241
Cognitive	195	188	243	418
Writing				
Background	37	61	61	61
Cognitive	6	6	6	12
Mathematics				
Background	47	89	147	220
Cognitive	241	327	412	665
Science				
Background	31	54	83	99
Cognitive	222	284	327	560
Excluded Student Questionnaire	67	67	67	67
Teacher Questionnaire	80	233	0	235
School Questionnaire	147	145	148	255
Total Items*	1265	1666	1801	2943

* Because many items were used at more than one age class and/or in more than one group of background questions, the total number of items is not equal to the total number of distinct items.

Table 16-3

Number of Cognitive Items Administered
by Sample and Age Class

	Age Class			
	9	13	17	Total
Cross-sectional				
Reading	67	97	112	176
Reading - Answer sheet	68	0	0	68
Mathematics	109	137	145	282
Special Math	48	75	81	105
Science	112	146	150	255
Bridge to 1984				
Reading	105	108	96	193
Writing	6	6	6	12
Bridge to 1986 (age and grade)				
Mathematics	0	0	164	164
Science	0	0	124	124
Bridge to 1986 (age only)				
Reading	30	34	71	90
Mathematics	68	98	94	184
Science	63	83	82	180
Long-term Bridge				
Mathematics	49	61	67	114
Science	66	77	78	134
Total Cognitive Items*	663	804	987	1654

* Because many items were used at more than one age class and/or for more than one sample, the total number of cognitive items is not equal to the total number of distinct items used for the three age classes and across the samples. Item counts in this table are for all items presented, but not necessarily scaled. Therefore, the number of items may not agree with item counts reported in other chapters.

Table 16-4

Characteristics of Schools in Main NAEP (Cross-sectional) Samples

	Age/Grade			
	9/4	13/8	17/12	Total
Total original sample	652	569	464	1685
Cooperating	523	402	301	1226
Out-of-range or closed	48	49	47	144
No eligibles enrolled	12	49	38	99
District refused	29	27	32	88
School refused	40	42	46	128
Cooperation rate	88.6%	86.7%	81.3%	86.0%
Cooperating replacements for refusals	4	4	3	11
Totals				
Cooperating schools	527	406	304	1237
Completing questionnaires	490	367	266	1123
Region				
Northeast	121	98	68	287
Southeast	109	91	88	288
Central	139	105	60	304
West	158	112	88	358
School type				
Public	327	216	200	743
Private	94	79	74	247
Catholic	106	111	30	247
Bureau of Indian Affairs	0	0	0	0
Department of Defense	0	0	0	0
Size and type of community				
Rural	48	44	42	134
Disadvantaged urban	50	40	32	122
Advantaged urban	76	67	40	183
Big city	63	47	18	128
Fringe	49	39	27	115
Medium city	86	53	45	184
Small place	155	116	100	371

Table 16-4 (continued)

Characteristics of Schools in Main NAEP (Cross-sectional) Samples

	Age/Grade			
	9/4	13/8	17/12	Total
Number of teachers				
Unclassified	41	47	40	128
1 - 4	18	14	7	39
5 - 9	65	43	14	122
10 - 19	153	106	49	308
20 - 49	223	138	86	447
50 - 74	26	43	44	113
75 - 99	1	12	27	40
100 +	0	3	37	40
Number of students				
Unclassified	40	47	43	130
1 - 99	29	27	22	78
100 - 299	176	117	58	351
300 - 499	140	83	34	257
500 - 749	88	55	34	177
750 - 999	40	40	30	110
1000 - 1499	13	28	33	74
1500 +	1	9	50	60

Table 16-5

Characteristics of Schools in NAEP Bridge Samples

	Age/Grade			
	9/4	13/8	17/12	Total
Total original sample	350	374	395	1119
Cooperating	283	282	292	857
Out-of-range or closed	15	7	22	44
No eligibles enrolled	12	50	9	71
District refused	23	19	31	73
School refused	17	16	41	74
Cooperation rate	88.1%	90.5%	80.7%	86.3%
Cooperating replacements for refusals			5	0 0 5
Totals				
Cooperating schools	288	282	292	862
Completing questionnaires	263	268	261	792
Region				
Northeast	65	58	60	183
Southeast	62	75	78	215
Central	76	69	75	220
West	85	80	79	244
School type				
Public	230	208	258	696
Private	23	30	23	76
Catholic	34	44	11	89
Bureau of Indian Affairs	1	0	0	1
Department of Defense	0	0	0	0
Size and type of community				
Rural	32	29	46	107
Disadvantaged urban	27	24	24	75
Advantaged urban	36	42	29	107
Big city	23	23	19	65
Fringe	25	24	28	77
Medium city	40	36	33	109
Small place	105	104	113	322

Table 16-5 (continued)

Characteristics of Schools in NAEP Bridge Samples

	Age/Grade			
	9/4	13/8	17/12	Total
Number of teachers				
Unclassified	28	17	35	80
1 - 4	4	3	3	10
5 - 9	27	22	4	53
10 - 19	81	53	37	171
20 - 49	136	130	95	361
50 - 74	10	42	52	104
75 - 99	1	9	29	39
100 +	1	6	37	44
Number of students				
Unclassified	32	18	36	86
1 - 99	12	11	7	30
100 - 299	69	66	42	177
300 - 499	77	52	43	172
500 - 749	69	58	31	158
750 - 999	21	37	32	90
1000 - 1499	6	36	43	85
1500 +	2	4	58	64

Table 16-6

Numbers of Responses to Teacher Questionnaires and Students Matched with Teacher Data*

	Number of Teachers Responding	Number of Students with No Match	Partial Match	Complete Match
Grade 4 Mathematics Questionnaire	393			
Main focused-BIB sample		627	404	5436
Special mathematics sample		215	121	2099
Grade 8 Mathematics Questionnaire	597			
Main focused-BIB sample		851	714	4908
Special mathematics sample		363	309	1743
Grade 8 Science Questionnaire	510	945	787	4799

* See section 9.2.5 in Chapter 9 for a discussion of student/teacher match rates.

Table 16-7

Number of Assessment Sessions by Sample, Type of Session, and Age Class

	Age Class			
	9	13	17	Total
Main Sample				
Regular	1339	1025	908	3272
Makeup	5	26	133	164
Bridge to 1984				
Regular	244	237	234	715
Makeup	1	8	30	39
Bridge to 1986 (age only)				
Regular	367	339	222	928
Makeup	4	1	37	42
Bridge to 1986 (age and grade)				
Regular	0	0	285	285
Makeup	0	0	39	39
Long-term Bridge				
Regular	242	237	222	701
Makeup	2	1	28	31
Total				
Regular	2192	1838	1871	5901
Makeup	12	36	267	315

Table 16-8

Number of Students Assessed and Excluded by Sample and Age Class

	Age Class 9	13	17	Total
Assessed				
Main and Special	32490	29250	28341	90081
Bridge to 1984	5926	6233	5614	17773
Bridge to 1986--age only	6235	6649	4411	17295
Bridge to 1986--grade/age	0	0	8338	8338
Long-term Trend	4134	4455	4402	12991
Excluded				
Main NAEP	2332	1950	1446	5728
Bridges	1116	1095	1239	3450
Total	52233	49632	53791	155656

Table 16-9

Numbers of Students in Main Sample
by Type of Eligibility and Subgroup Classification
Age 9/Grade 4

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Total	23250	24257	15017	32490
Sex				
Male	11689	12309	7045	16953
Female	11561	11948	7972	15537
Race/Ethnicity				
White	13822	14232	9102	18952
Black	3655	4152	2379	5428
Hispanic	4395	4413	2541	6267
Other	1378	1460	995	1843
Region				
Northeast	4673	4985	3455	6203
Southeast	5726	6239	3657	8308
Central	5392	5614	3221	7785
West	7459	7419	4684	10194
Parents' Education				
Less than high school	1070	1148	603	1615
High school	3067	3490	1997	4560
Greater than high school	1721	1960	1231	2450
Graduated college	8369	9034	5754	11649
Unknown	8849	8497	5346	12000
Size and Type of Community				
Rural	1818	1820	1032	2606
Disadvantaged urban	2866	3059	1911	4014
Advantaged urban	2845	2925	1987	3783
Big city	2787	3002	1987	3802
Fringe	2459	2540	1716	3283
Medium city	3841	4232	2510	5563
Small place	6634	6679	3874	9439

Table 16-10

Numbers of Students in Main Sample
by Type of Eligibility and Subgroup Classification
Age 13/Grade 8

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Total	20984	21929	13663	29250
Sex				
Male	10221	10877	6115	14983
Female	10763	11052	7548	14267
Race/Ethnicity				
White	13577	14142	9238	18481
Black	2874	3202	1734	4342
Hispanic	3286	3261	1798	4749
Other	1247	1324	893	1678
Region				
Northeast	4866	5169	3634	6401
Southeast	5110	5418	3237	7291
Central	4408	4629	2799	6238
West	6600	6713	3993	9320
Parents' Education				
Less than high school	1539	1908	843	2604
High school	4973	5229	3023	7179
Greater than high school	3652	3977	2640	4989
Graduated college	8758	8933	6152	11539
Unknown	2008	1842	982	2868
Size and Type of Community				
Rural	1903	2100	1134	2869
Disadvantaged urban	2820	2802	1691	3931
Advantaged urban	2764	2735	2001	3498
Big city	2555	2583	1701	3437
Fringe	2345	2616	1718	3243
Medium city	2706	2769	1753	3722
Small place	5891	6324	3665	8550

Table 16-11

Numbers of Students in Main Sample
by Type of Eligibility and Subgroup Classification
Age 17/Grade 12

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Total	21614	21277	14550	28341
Sex				
Male	10655	10311	6493	14473
Female	10959	10966	8057	13868
Race/Ethnicity				
White	15143	15095	10677	19561
Black	3046	2936	1839	4143
Hispanic	2375	2216	1356	3235
Other	1050	1030	678	1402
Region				
Northeast	5200	5453	3910	6743
Southeast	6244	5898	4157	7985
Central	4242	4261	2742	5761
West	5928	5665	3741	7852
Parents' Education				
Less than high school	1798	1694	949	2543
High school	5038	4805	3105	6738
Greater than high school	5347	5341	3821	6867
Graduated college	8842	8960	6414	11388
Unknown	536	428	229	735
Size and Type of Community				
Rural	2261	2212	1481	2992
Disadvantaged urban	2884	2748	1779	3853
Advantaged urban	2538	2614	1900	3252
Big city	1619	1636	1183	2072
Fringe	2733	2680	2017	3396
Medium city	2889	2769	1922	3736
Small place	6690	6618	4268	9040

Table 16-12

Numbers of Students in Bridge to 1984 Sample
by Type of Eligibility and Subgroup Classification
Age 9/Grade 4

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Total	4268	4367	2709	5926
Sex				
Male	2133	2188	1269	3052
Female	2135	2179	1440	2874
Race/Ethnicity				
White	2659	2757	1725	3691
Black	636	686	386	936
Hispanic	747	696	439	1004
Other	226	228	159	295
Region				
Northeast	914	904	637	1181
Southeast	1089	1124	641	1572
Central	948	951	564	1335
West	1317	1388	867	1838
Parents' Education				
Less than high school	213	228	118	323
High school	716	791	469	1038
Greater than high school	223	214	147	290
Graduated college	1611	1721	1103	2229
Unknown	1497	1407	870	2034
Size and Type of Community				
Rural	338	377	203	512
Disadvantaged urban	513	508	324	697
Advantaged urban	471	470	344	597
Big city	326	318	222	422
Fringe	440	444	283	601
Medium city	628	608	370	866
Small place	1552	1642	963	2231

Table 16-13

Numbers of Students in Bridge to 1984 Sample
by Type of Eligibility and Subgroup Classification
Age 13/Grade 8

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Total	4609	4665	3041	6233
Sex				
Male	2304	2346	1411	3239
Female	2305	2319	1630	2994
Race/Ethnicity				
White	3184	3237	2162	4259
Black	559	566	347	778
Hispanic	624	616	361	879
Other	242	246	171	317
Region				
Northeast	1030	985	749	1266
Southeast	1086	1186	738	1534
Central	1063	1041	655	1449
West	1430	1453	899	1984
Parents' Education				
Less than high school	340	342	167	515
High school	1382	1449	914	1917
Greater than high school	522	533	359	696
Graduated college	1878	1887	1337	2428
Unknown	474	441	257	658
Size and Type of Community				
Rural	321	369	197	493
Disadvantaged urban	491	434	283	642
Advantaged urban	654	679	491	842
Big city	278	306	210	374
Fringe	551	531	417	665
Medium city	513	538	326	725
Small place	1801	1808	1117	2492

Table 16-14

Numbers of Students in Bridge to 1984 Sample
by Type of Eligibility and Subgroup Classification
Age 17/Grade 11

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Total	4383	4385	3154	5614
Sex				
Male	2188	2206	1491	2903
Female	2195	2179	1663	2711
Race/Ethnicity				
White	3247	3222	2448	4021
Black	613	601	375	839
Hispanic	314	342	198	458
Other	209	220	133	296
Region				
Northeast	1000	989	693	1296
Southeast	1157	1143	811	1489
Central	1138	1120	854	1404
West	1088	1133	796	1425
Parents' Education				
Less than high school	361	345	204	502
High school	1294	1285	897	1682
Greater than high school	816	826	609	1033
Graduated college	1775	1807	1378	2204
Unknown	124	113	60	177
Size and Type of Community				
Rural	552	554	400	706
Disadvantaged urban	336	358	207	487
Advantaged urban	487	504	370	621
Big city	247	257	176	328
Fringe	507	501	364	644
Medium city	669	633	460	842
Small place	1585	1578	1177	1986

Table 16-15

Numbers of Students in Bridge to 1986 Sample (Age only)
by Type of Eligibility and Subgroup Classification
Age 9

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Total	6235	4016	4016	6235
Sex				
Male	3101	1870	1870	3101
Female	3134	2146	2146	3134
Race/Ethnicity				
White	3897	2564	2564	3897
Black	900	564	564	900
Hispanic	1170	702	702	1170
Other	268	186	186	268
Region				
Northeast	1336	984	984	1336
Southeast	1485	869	869	1485
Central	1513	909	909	1513
West	1901	1254	1254	1901
Parents' Education				
Less than high school	306	156	156	306
High school	1003	623	623	1003
Greater than high school	466	324	324	466
Graduated college	2462	1686	1686	2462
Unknown	1991	1223	1223	1991
Size and Type of Community				
Rural	401	242	242	401
Disadvantaged urban	637	432	432	637
Advantaged urban	749	558	558	749
Big city	507	346	346	507
Fringe	572	360	360	572
Medium city	873	538	538	873
Small place	2496	1540	1540	2496

Table 16-16

Numbers of Students in Bridge to 1986 Sample (Age only)
by Type of Eligibility and Subgroup Classification
Age 13

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Total	6649	4330	4330	6649
Sex				
Male	3299	1985	1985	3299
Female	3350	2345	2345	3350
Race/Ethnicity				
White	4737	3167	3167	4737
Black	688	417	417	688
Hispanic	859	474	474	859
Other	365	272	272	365
Region				
Northeast	1334	937	937	1334
Southeast	1738	1154	1154	1738
Central	1529	943	943	1529
West	2048	1296	1296	2048
Parents' Education				
Less than high school	515	277	277	515
High school	1743	1065	1065	1743
Greater than high school	1145	847	847	1145
Graduated college	2715	1882	1882	2715
Unknown	516	251	251	516
Size and Type of Community				
Rural	557	325	325	557
Disadvantaged urban	642	380	380	642
Advantaged urban	667	475	475	667
Big city	559	415	415	559
Fringe	726	504	504	726
Medium city	785	511	511	785
Small place	2713	1720	1720	2713

Table 16-17

Numbers of Students in Bridge to 1986 Sample (Age only)
by Type of Eligibility and Subgroup Classification
Age 17

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Total	4411	3173	3173	4411
Sex				
Male	2154	1475	1475	2154
Female	2257	1698	1698	2257
Race/Ethnicity				
White	3318	2477	2477	3318
Black	587	376	376	587
Hispanic	329	205	205	329
Other	177	115	115	177
Region				
Northeast	985	671	671	985
Southeast	1170	795	795	1170
Central	1186	919	919	1186
West	1070	788	788	1070
Parents' Education				
Less than high school	339	198	198	339
High school	1152	782	782	1152
Greater than high school	1047	813	813	1047
Graduated college	1755	1318	1318	1755
Unknown	112	60	60	112
Size and Type of Community				
Rural	521	361	361	521
Disadvantaged urban	360	214	214	360
Advantaged urban	500	387	387	500
Big city	282	204	204	282
Fringe	522	377	377	522
Medium city	587	416	416	587
Small place	1639	1214	1214	1639

Table 16-18

Numbers of Students in Bridge to 1986 Sample (Age and Grade)
by Type of Eligibility and Subgroup Classification
Age 17/Grade 11

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Total	6557	6426	4645	8338
Sex				
Male	3180	3110	2129	4161
Female	3377	3316	2516	4177
Race/Ethnicity				
White	4841	4743	3581	6003
Black	922	864	569	1217
Hispanic	484	502	286	700
Other	310	317	209	418
Region				
Northeast	1395	1385	979	1801
Southeast	1704	1643	1121	2226
Central	1796	1726	1335	2187
West	1662	1672	1210	2124
Parents' Education				
Less than high school	572	523	327	768
High school	1703	1626	1142	2187
Greater than high school	1588	1580	1184	1984
Graduated college	2486	2500	1888	3098
Unknown	192	183	98	277
Size and Type of Community				
Rural	908	867	615	1160
Disadvantaged urban	552	536	318	770
Advantaged urban	742	768	575	935
Big city	452	448	308	592
Fringe	707	685	496	896
Medium city	954	945	679	1220
Small place	2242	2177	1654	2765

Table 16-19

Numbers of Students in Long-term Trend Sample
by Type of Eligibility and Subgroup Classification
Age 9

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Total	4134	2661	2661	4134
Sex				
Male	2060	1219	1219	2060
Female	2074	1442	1442	2074
Race/Ethnicity				
White	2523	1631	1631	2523
Black	563	371	371	563
Hispanic	804	497	497	804
Other	244	162	162	244
Region				
Northeast	844	615	615	844
Southeast	960	573	573	960
Central	939	570	570	939
West	1391	903	903	1391
Parents' Education				
Less than high school	192	106	106	192
High school	608	400	400	608
Greater than high school	309	211	211	309
Graduated college	1627	1100	1100	1627
Unknown	1392	841	841	1392
Size and Type of Community				
Rural	302	188	188	302
Disadvantaged urban	429	276	276	429
Advantaged urban	462	337	337	462
Big city	368	263	263	368
Fringe	442	287	287	442
Medium city	631	371	371	631
Small place	1500	939	939	1500

Table 16-20

Numbers of Students in Long-term Trend Sample
by Type of Eligibility and Subgroup Classification
Age 13

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Total	4455	2854	2854	4455
Sex				
Male	2167	1270	1270	2167
Female	2288	1584	1584	2288
Race/Ethnicity				
White	3067	2059	2059	3067
Black	563	295	295	563
Hispanic	562	321	321	562
Other	263	179	179	263
Region				
Northeast	831	596	596	831
Southeast	1233	775	775	1233
Central	1028	611	611	1028
West	1363	872	872	1363
Parents' Education				
Less than high school	336	188	188	336
High school	1205	711	711	1205
Greater than high school	784	549	549	784
Graduated college	1769	1213	1213	1769
Unknown	344	180	180	344
Size and Type of Community				
Rural	470	272	272	470
Disadvantaged urban	468	280	280	468
Advantaged urban	577	409	409	577
Big city	344	254	254	344
Fringe	409	286	286	409
Medium city	485	310	310	485
Small place	1702	1043	1043	1702

Table 16-21

Numbers of Students in Long-term Trend Sample
by Type of Eligibility and Subgroup Classification
Age 17

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Total	4402	3149	3149	4402
Sex				
Male	2149	1447	1447	2149
Female	2253	1702	1702	2253
Race/Ethnicity				
White	3291	2492	2492	3291
Black	600	367	367	600
Hispanic	328	180	180	328
Other	183	110	110	183
Region				
Northeast	1021	706	706	1021
Southeast	1117	781	781	1117
Central	1123	846	846	1123
West	1141	816	816	1141
Parents' Education				
Less than high school	355	204	204	355
High school	1118	722	722	1118
Greater than high school	1073	803	803	1073
Graduated college	1766	1370	1370	1766
Unknown	63	30	30	63
Size and Type of Community				
Rural	561	395	395	561
Disadvantaged urban	355	211	211	355
Advantaged urban	460	343	343	460
Big city	285	201	201	285
Fringe	509	374	374	509
Medium city	592	413	413	592
Small place	1640	1212	1212	1640

Table 16-22

Numbers of Excluded Students in Main Sample
by Type of Eligibility and Subgroup Classification*
Age 9/Grade 4

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Total	1456	1425	549	2332
Sex				
Male	907	912	330	1489
Female	549	513	219	843
Race/Ethnicity				
White	518	507	156	869
Black	238	230	68	400
Hispanic	494	488	235	747
Other	206	200	90	316
Region				
Northeast	308	284	131	461
Southeast	223	242	44	421
Central	194	190	39	345
West	731	709	335	1105
Size and Type of Community				
Rural	103	56	16	143
Disadvantaged urban	419	394	185	628
Advantaged urban	71	52	28	95
Big city	126	144	58	212
Fringe	225	244	111	358
Medium city	199	206	53	352
Small place	313	329	98	544

* Data on parents' education were not collected for excluded students.

Table 16-23

Numbers of Excluded Students in Main Sample
by Type of Eligibility and Subgroup Classification*
Age 13/Grade 8

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Total	1142	1183	375	1950
Sex				
Male	717	750	218	1249
Female	425	433	157	701
Race/Ethnicity				
White	446	484	135	795
Black	205	199	50	354
Hispanic	290	323	115	498
Other	201	177	75	303
Region				
Northeast	217	222	86	353
Southeast	173	148	26	295
Central	263	322	86	499
West	489	491	177	803
Size and Type of Community				
Rural	83	86	18	151
Disadvantaged urban	280	278	103	455
Advantaged urban	48	43	16	75
Big city	171	166	70	267
Fringe	121	119	51	189
Medium city	193	205	54	344
Small place	246	286	63	469

* Data on parents' education were not collected for excluded students.

Table 16-24

Numbers of Excluded Students in Main Sample
by Type of Eligibility and Subgroup Classification*
Age 17/Grade 12

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Total	920	728	202	1446
Sex				
Male	593	442	103	932
Female	327	286	99	514
Race/Ethnicity				
White	400	370	94	676
Black	178	135	25	288
Hispanic	208	125	49	284
Other	134	98	34	198
Region				
Northeast	143	112	34	221
Southeast	242	188	55	375
Central	150	157	39	268
West	385	271	74	582
Size and Type of Community				
Rural	96	103	20	179
Disadvantaged urban	267	156	47	376
Advantaged urban	38	31	5	64
Big city	67	51	18	100
Fringe	84	60	24	120
Medium city	146	121	32	235
Small place	222	206	56	372

* Data on parents' education were not collected for excluded students.

Table 16-25

Numbers of Excluded Students in Bridge Samples
by Type of Eligibility and Subgroup Classification*
Age 9/Grade 4

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Total	949	571	404	1116
Sex				
Male	585	352	234	703
Female	364	219	170	413
Race/Ethnicity				
White	382	196	125	453
Black	110	70	36	144
Hispanic	311	214	176	349
Other	73	49	41	81
Region				
Northeast	247	144	110	281
Southeast	116	87	35	168
Central	171	81	50	202
West	415	259	209	465
Size and Type of Community				
Rural	29	24	12	41
Disadvantaged urban	185	117	90	212
Advantaged urban	95	46	37	104
Big city	111	89	81	119
Fringe	95	54	42	107
Medium city	136	74	47	163
Small place	298	167	95	370

* Data on parents' education were not collected for excluded students.

Table 16-26

Numbers of Excluded Students in Bridge Samples by Type of Eligibility and Subgroup Classification*
Age 13/Grade 8

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Total	923	478	306	1095
Sex				
Male	572	282	169	685
Female	351	196	137	410
Race/Ethnicity				
White	409	204	117	496
Black	128	51	27	152
Hispanic	254	164	117	301
Other	132	59	45	146
Region				
Northeast	181	89	64	206
Southeast	232	109	70	271
Central	157	63	24	196
West	353	217	148	422
Size and Type of Community				
Rural	72	41	25	88
Disadvantaged urban	248	139	110	277
Advantaged urban	77	30	20	87
Big city	91	60	45	106
Fringe	72	45	24	93
Medium city	99	46	18	127
Small place	264	117	64	317

* Data on parents' education were not collected for excluded students.

Table 16-27

Numbers of Excluded Students in Bridge Samples
by Type of Eligibility and Subgroup Classification*
Age 17/Grade 11

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Total	928	585	274	1239
Sex				
Male	604	385	168	821
Female	324	200	106	418
Race/Ethnicity				
White	482	290	147	625
Black	144	109	41	212
Hispanic	150	75	31	194
Other	152	111	55	208
Region				
Northeast	142	95	38	199
Southeast	234	153	70	317
Central	211	132	66	277
West	341	205	100	446
Size and Type of Community				
Rural	108	58	32	134
Disadvantaged urban	138	65	27	176
Advantaged urban	69	40	15	94
Big city	69	52	21	100
Fringe	110	70	35	145
Medium city	127	84	31	180
Small place	307	216	113	410

* Data on parents' education were not collected for excluded students.

Table 16-28

Number of Students by Sample and Age Class

	Age Class 9		Age Class 13		Age Class 17	
	Total	Sum of Weights	Total	Sum of Weights	Total	Sum of Weights
Main Assessment						
Winter and Spring combined	32490	4825560	29250	4431116	28341	3843752
Winter only	16818	4798658	15898	4414103	14644	3830471
Spring only	15672	4852462	13352	4448130	13697	3857032
Bridges						
Bridge to 1984	5926	4828562	6233	4389803	5614	4087595
Bridge to 1986--age only	6235	3637675	6649	3105980	4411	3002755
Bridge to 1986--grade/age	0	0	0	0	8338	4084061
Long-term trend	4134	3599721	4455	3116104	4402	3000558
Excluded Students						
Main	2332	650330	1950	587313	1446	344476
Bridge	1116	338645	1095	297717	1239	237316

Table 16-29

Number of Students Assessed in Main Assessment
Age 9/Grade 4 (Booklets 1-28)

	Grade < 4	Grade = 4	Grade > 4	Total
AGE < 9				
Unweighted N	0	139	0	139
Estimated population size	0	18151	0	18151
Standard error	0	1987	0	1987
Coefficient of variation*	0.00	10.95	0.00	10.95
AGE = 9				
Unweighted N	8157	15017	76	23250
Estimated population size	1432200	2009536	10648	3452384
Standard error	11087	27112	1877	29471
Coefficient of variation*	0.77	1.35	17.63	0.85
AGE > 9				
Unweighted N	0	9101	0	9101
Estimated population size	0	1355025	0	1355025
Standard error	0	24237	0	24237
Coefficient of variation*	0.00	1.79	0.00	1.79
AGE TOTAL				
Unweighted N	8157	24257	76	32490
Estimated population size	1432200	3382712	10648	4825560
Standard error	11087	11505	1877	17869
Coefficient of variation*	0.77	0.34	17.63	0.37

* The coefficient of variation of the estimated population size is defined as 100 times its standard error divided by the estimated population size.

Table 16-30

Number of Students Assessed in Main Assessment
Age 13/Grade 8 (Booklets 1-28)

	Grade < 8	Grade = 8	Grade > 8	Total
AGE < 13				
Unweighted N	0	181	0	181
Estimated population size	0	23174	0	23174
Standard error	0	2438	0	2438
Coefficient of variation*	0.00	10.52	0.00	10.52
AGE = 13				
Unweighted N	7264	13663	57	20984
Estimated population size	1338568	1786948	20134	3145651
Standard error	11941	24921	8980	25730
Coefficient of variation*	0.89	1.39	44.60	0.82
AGE > 13				
Unweighted N	0	8085	0	8085
Estimated population size	0	1262291	0	1262291
Standard error	0	26789	0	26789
Coefficient of variation*	0.00	2.12	0.00	2.12
AGE TOTAL				
Unweighted N	7264	21929	57	29250
Estimated population size	1338568	3072414	20134	4431116
Standard error	11941	10624	8980	15606
Coefficient of variation*	0.89	0.35	44.60	0.35

* The coefficient of variation of the estimated population size is defined as 100 times its standard error divided by the estimated population size.

Table 16-31

Number of Students Assessed in Main Assessment
Age 17/Grade 12 (Booklets 1-28)

	Grade			
	< 12	= 12	> 12	Total
AGE < 17				
Unweighted N	0	276	0	276
Estimated population size	0	32109	0	32109
Standard error	0	3668	0	3668
Coefficient of variation*	0.00	11.42	0.00	11.42
AGE = 17				
Unweighted N	7064	14550	0	21614
Estimated population size	984467	1889011	0	2873478
Standard error	5759	13743	0	15061
Coefficient of variation*	0.58	0.73	0.00	0.52
AGE > 17				
Unweighted N	0	6451	0	6451
Estimated population size	0	938165	0	938165
Standard error	0	29896	0	29896
Coefficient of variation*	0.00	3.19	0.00	3.19
AGE TOTAL				
Unweighted N	7064	21277	0	28341
Estimated population size	984467	2859285	0	3843752
Standard error	5759	21116	0	21630
Coefficient of variation*	0.58	0.74	0.00	0.56

* The coefficient of variation of the estimated population size is defined as 100 times its standard error divided by the estimated population size.

Table 16-32

Number of Students Assessed in Bridge to 1984
Age 9/Grade 4 (Booklets 51-56)

	Grade < 4	Grade = 4	Grade > 4	Total
AGE < 9				
Unweighted N	0	13	0	13
Estimated population size	0	10808	0	10808
Standard error	0	2449	0	2449
Coefficient of variation*	0.00	22.66	0.00	22.66
AGE = 9				
Unweighted N	1549	2709	10	4268
Estimated population size	1457675	2001815	8095	3467586
Standard error	12702	36416	2565	38225
Coefficient of variation*	0.87	1.82	31.69	1.10
AGE > 9				
Unweighted N	0	1645	0	1645
Estimated population size	0	1350168	0	1350168
Standard error	0	33784	0	33784
Coefficient of variation*	0.00	2.50	0.00	2.50
AGE TOTAL				
Unweighted N	1549	4367	10	5926
Estimated population size	1457675	3362791	8095	4828562
Standard error	12702	18271	2565	26983
Coefficient of variation*	0.87	0.54	31.69	0.56

* The coefficient of variation of the estimated population size is defined as 100 times its standard error divided by the estimated population size.

Table 16-33

Number of Students Assessed in Bridge to 1984
Age 13/Grade 8 (Booklets 51-56)

	Grade < 8	Grade = 8	Grade > 8	Total
AGE < 13				
Unweighted N	0	27	0	27
Estimated population size	0	22356	0	22356
Standard error	0	5190	0	5190
Coefficient of variation*	0.00	23.22	0.00	23.22
AGE = 13				
Unweighted N	1557	3041	11	4609
Estimated population size	1226825	1870778	15361	3112964
Standard error	15893	6644	8065	15907
Coefficient of variation*	1.30	0.36	52.50	0.51
AGE > 13				
Unweighted N	0	1597	0	1597
Estimated population size	0	1254483	0	1254483
Standard error	0	15373	0	15373
Coefficient of variation*	0.00	1.23	0.00	1.23
AGE TOTAL				
Unweighted N	1557	4665	11	6233
Estimated population size	1226825	3147617	15361	4389803
Standard error	15893	17211	8065	26114
Coefficient of variation*	1.30	0.55	52.50	0.59

* The coefficient of variation of the estimated population size is defined as 100 times its standard error divided by the estimated population size.

Table 16-34

Number of Students Assessed in Bridge to 1984
Age 17/Grade 11 (Booklets 51-56)

	Grade < 11	Grade = 11	Grade > 11	Total
AGE < 17				
Unweighted N	0	391	0	391
Estimated population size	0	341468	0	341468
Standard error	0	21374	0	21374
Coefficient of variation*	0.00	6.26	0.00	6.26
AGE = 17				
Unweighted N	899	3154	330	4383
Estimated population size	780211	1945753	260040	2986005
Standard error	19304	3221	17451	11167
Coefficient of variation*	2.47	0.17	6.71	0.37
AGE > 17				
Unweighted N	0	840	0	840
Estimated population size	0	760122	0	760122
Standard error	0	21260	0	21260
Coefficient of variation*	0.00	2.80	0.00	2.80
AGE TOTAL				
Unweighted N	899	4385	330	5614
Estimated population size	780211	3047343	260040	4087595
Standard error	19304	9730	17451	15320
Coefficient of variation*	2.47	0.32	6.71	0.37

* The coefficient of variation of the estimated population size is defined as 100 times its standard error divided by the estimated population size.

Table 16-35

Number of Students Assessed in Bridge to 1986
Age Only

	Grade < 4	Grade = 4	Grade > 4	Total
AGE 9 (Booklets 91-93)				
Unweighted N	2202	4016	17	6235
Estimated population size	1267589	2358141	11944	3637675
Standard error	44073	64746	3413	34148
Coefficient of variation*	3.48	2.75	28.58	0.94
	< 8	= 8	> 8	Total
AGE 13 (Booklets 91-93)				
Unweighted N	2299	4330	20	6649
Estimated population size	1122502	1967661	15817	3105980
Standard error	41978	43688	7486	15307
Coefficient of variation*	3.74	2.22	47.33	0.49
	< 11	= 11	> 11	Total
AGE 17 (Booklets 84-85)				
Unweighted N	888	3173	350	4411
Estimated population size	651244	2108602	242909	3002755
Standard error	29106	29048	18102	7281
Coefficient of variation*	4.47	1.38	7.45	0.24

* The coefficient of variation of the estimated population size is defined as 100 times its standard error divided by the estimated population size.

Table 16-36

Number of Students Assessed in Bridge to 1986
Age 17/Grade 11 (Booklets 61-66)

	Grade < 11	Grade = 11	Grade > 11	Total
AGE < 17				
Unweighted N	0	566	0	566
Estimated population size	0	345811	0	345811
Standard error	0	17110	0	17110
Coefficient of variation*	0.00	4.95	0.00	4.95
AGE = 17				
Unweighted N	1423	4645	489	6557
Estimated population size	781322	1945083	262646	2989051
Standard error	19048	3058	18021	8345
Coefficient of variation*	2.44	0.16	6.86	0.28
AGE > 17				
Unweighted N	0	1215	0	1215
Estimated population size	0	749199	0	749199
Standard error	0	19652	0	19652
Coefficient of variation*	0.00	2.62	0.00	2.62
AGE TOTAL				
Unweighted N	1423	6426	489	8338
Estimated population size	781322	3040093	262646	4084061
Standard error	19048	10357	18021	13732
Coefficient of variation*	2.44	0.34	6.86	0.34

* The coefficient of variation of the estimated population size is defined as 100 times its standard error divided by the estimated population size.

Table 16-37

Number of Students Assessed in Long-term Trend
Age Only (Booklets 94-95)

	Grade < 4	Grade = 4	Grade > 4	Total
AGE 9				
Unweighted N	1460	2661	13	4134
Estimated population size	1290065	2299759	9898	3599721
Standard error	54587	74814	3108	35162
Coefficient of variation*	4.23	3.25	31.40	0.98
	< 8	= 8	> 8	Total
AGE 13				
Unweighted N	1582	2854	19	4455
Estimated population size	1176978	1904527	34599	3116104
Standard error	49725	52031	15888	14706
Coefficient of variation*	4.22	2.73	45.92	0.47
	< 11	= 11	> 11	Total
AGE 17				
Unweighted N	924	3149	329	4402
Estimated population size	660421	2120949	219187	3000558
Standard error	23670	25579	17909	8916
Coefficient of variation*	3.58	1.21	8.17	0.30

* The coefficient of variation of the estimated population size is defined as 100 times its standard error divided by the estimated population size.

Table 16-38

Weighted Percentage of Students in Main Sample
by Type of Eligibility and Subgroup Classification
Age 9/Grade 4

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Sex				
Male	50.48	50.79	46.11	52.52
Female	49.52	49.21	53.89	47.48
Race/Ethnicity				
White	70.72	70.13	71.98	69.78
Black	14.62	15.27	13.88	15.38
Hispanic	10.36	10.77	9.73	10.91
Other	4.31	3.84	4.41	3.93
Region				
Northeast	21.18	21.69	24.44	20.18
Southeast	24.29	24.18	22.95	24.77
Central	25.54	25.88	23.81	26.50
West	29.00	28.25	28.80	28.55
Parents' Education				
Less than high school	4.72	5.08	4.03	5.26
High school	13.90	15.57	14.35	14.88
Greater than high school	7.45	8.45	8.67	7.64
Graduated college	34.83	35.89	37.23	34.57
Unknown	38.33	34.52	35.17	36.97
Size and Type of Community				
Rural	11.10	10.82	9.96	11.38
Disadvantaged urban	9.45	9.55	9.69	9.42
Advantaged urban	11.49	11.39	12.55	10.98
Big city	8.88	9.25	10.24	8.57
Fringe	11.43	11.60	12.71	11.01
Medium city	12.97	14.05	13.50	13.51
Small place	34.69	33.33	31.36	35.12
Estimated Total Population	3452384	3382712	2009536	4825560

Table 16-39

Weighted Percentage of Students in Main Sample
by Type of Eligibility and Subgroup Classification
Age 13/Grade 8

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Sex				
Male	49.33	49.87	44.22	51.76
Female	50.67	50.13	55.78	48.24
Race/Ethnicity				
White	69.96	70.54	73.37	68.99
Black	14.72	15.11	13.34	15.55
Hispanic	11.04	10.03	8.69	11.29
Other	4.28	4.32	4.60	4.18
Region				
Northeast	20.99	20.98	23.64	19.91
Southeast	24.07	24.28	22.75	24.75
Central	25.33	24.84	24.51	25.32
West	29.61	29.90	29.09	30.02
Parents' Education				
Less than high school	7.50	8.87	5.96	9.07
High school	25.20	25.02	23.02	25.96
Greater than high school	17.36	18.40	19.84	17.08
Graduated college	40.56	39.83	44.67	38.40
Unknown	9.11	7.71	6.34	9.25
Size and Type of Community				
Rural	11.22	11.51	10.24	11.81
Disadvantaged urban	10.38	9.60	9.29	10.28
Advantaged urban	11.65	10.64	12.34	10.67
Big city	9.48	9.22	10.11	9.05
Fringe	10.21	11.73	12.61	10.30
Medium city	12.32	12.03	12.41	12.08
Small place	34.74	35.27	33.00	35.81
Estimated Total Population	3145651	3072414	1786948	4431116

Table 16-40

Weighted Percentage of Students in Main Sample
by Type of Eligibility and Subgroup Classification
Age 17/Grade 12

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Sex				
Male	49.15	48.61	44.48	51.04
Female	50.85	51.39	55.52	48.96
Race/Ethnicity				
White	71.59	73.49	75.76	70.96
Black	15.07	14.26	12.93	15.52
Hispanic	9.20	8.01	7.05	9.37
Other	4.13	4.24	4.26	4.15
Region				
Northeast	22.22	23.83	24.79	22.15
Southeast	23.47	20.70	21.38	22.43
Central	25.15	26.65	25.29	26.19
West	29.15	28.83	28.53	29.22
Parents' Education				
Less than high school	8.19	7.68	6.15	8.81
High school	24.54	23.86	22.28	25.14
Greater than high school	24.81	25.36	26.45	24.41
Graduated college	39.65	40.82	43.32	38.72
Unknown	2.54	2.07	1.59	2.65
Size and Type of Community				
Rural	11.34	11.49	11.24	11.49
Disadvantaged urban	12.73	12.11	11.52	12.86
Advantaged urban	9.03	9.62	10.14	8.92
Big city	5.17	5.44	5.76	5.08
Fringe	12.08	11.77	13.19	11.30
Medium city	12.89	12.29	12.19	12.79
Small place	36.76	37.28	35.95	37.55
Estimated Total Population	2873478	2859285	1889011	3843752

Table 16-41

Weighted Percentage of Students in Bridge to 1984 Sample
by Type of Eligibility and Subgroup Classification
Age 9/Grade 4

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Sex				
Male	50.84	50.83	46.85	52.49
Female	49.16	49.17	53.15	47.51
Race/Ethnicity				
White	70.31	70.37	71.05	70.05
Black	14.85	15.21	13.99	15.46
Hispanic	10.57	10.67	10.49	10.67
Other	4.27	3.75	4.47	3.83
Region				
Northeast	22.33	22.02	25.45	20.82
Southeast	24.22	24.22	22.08	25.10
Central	24.99	24.94	23.53	25.56
West	28.47	28.82	28.94	28.52
Parents' Education				
Less than high school	4.98	5.31	4.24	5.52
High school	17.08	18.63	17.48	17.99
Greater than high school	5.19	4.72	5.40	4.78
Graduated college	37.13	39.27	41.14	36.96
Unknown	35.49	31.93	31.69	34.59
Size and Type of Community				
Rural	8.85	10.10	8.28	9.95
Disadvantaged urban	10.19	9.91	10.34	9.94
Advantaged urban	10.60	10.53	12.66	9.69
Big city	7.23	6.98	7.91	6.77
Fringe	9.10	8.59	8.92	8.82
Medium city	15.59	13.97	13.74	15.23
Small place	38.45	39.92	38.16	39.59
Estimated Total Population	3467586	3362791	2001815	4828562

Table 16-42

Weighted Percentage of Students in Bridge to 1984 Sample
by Type of Eligibility and Subgroup Classification
Age 13/Grade 8

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Sex				
Male	50.24	50.46	45.64	52.35
Female	49.76	49.54	54.36	47.65
Race/Ethnicity				
White	70.11	70.22	74.03	68.52
Black	14.89	15.10	12.95	15.87
Hispanic	10.80	10.32	8.79	11.32
Other	4.19	4.35	4.24	4.29
Region				
Northeast	23.20	22.82	24.95	22.18
Southeast	23.30	24.58	22.11	24.72
Central	23.71	23.23	23.55	23.43
West	29.79	29.38	29.38	29.67
Parents' Education				
Less than high school	7.94	7.60	5.52	8.72
High school	30.81	32.70	31.45	31.90
Greater than high school	11.41	11.65	12.24	11.23
Graduated college	38.81	38.13	42.32	36.83
Unknown	10.73	9.62	8.27	10.98
Size and Type of Community				
Rural	8.34	10.04	7.62	9.87
Disadvantaged urban	9.94	8.64	8.72	9.53
Advantaged urban	11.48	11.42	13.19	10.71
Big city	6.11	6.64	7.14	6.06
Fringe	10.02	9.54	11.50	9.04
Medium city	11.91	11.77	11.01	12.19
Small place	42.19	41.95	40.82	42.60
Estimated Total Population	3112964	3147617	1870778	4389803

Table 16-43

Weighted Percentage of Students in Bridge to 1984 Sample
by Type of Eligibility and Subgroup Classification
Age 17/Grade 11

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Sex				
Male	50.25	50.92	46.78	52.40
Female	49.75	49.08	53.22	47.60
Race/Ethnicity				
White	71.40	71.24	75.60	69.28
Black	15.37	15.49	12.98	16.59
Hispanic	9.02	8.75	7.42	9.58
Other	4.22	4.52	3.99	4.55
Region				
Northeast	22.56	22.13	21.59	22.70
Southeast	23.18	23.36	22.36	23.71
Central	26.64	26.83	29.15	25.59
West	27.61	27.68	26.91	28.00
Parents' Education				
Less than high school	8.81	8.48	6.74	9.55
High school	29.83	29.54	28.37	30.30
Greater than high school	18.48	18.42	19.18	18.10
Graduated college	39.38	40.41	43.43	38.22
Unknown	3.17	2.96	2.10	3.52
Size and Type of Community				
Rural	12.61	12.86	12.82	12.70
Disadvantaged urban	8.36	8.89	7.02	9.39
Advantaged urban	10.38	10.91	11.10	10.43
Big city	6.80	7.13	6.81	7.05
Fringe	11.52	11.27	11.63	11.28
Medium city	15.13	14.11	14.43	14.71
Small place	35.19	34.83	36.19	34.45
Estimated Total Population	2986005	3047343	1945753	4087595

Table 16-44

Weighted Percentage of Students in Bridge to 1986 Sample (Age Only)
by Type of Eligibility and Subgroup Classification
Age 9

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Sex				
Male	49.43	46.44	46.44	49.43
Female	50.57	53.56	53.56	50.57
Race/Ethnicity				
White	70.19	71.98	71.98	70.19
Black	15.25	14.41	14.41	15.25
Hispanic	10.31	9.20	9.20	10.31
Other	4.26	4.41	4.41	4.26
Region				
Northeast	23.21	26.57	26.57	23.21
Southeast	24.12	22.26	22.26	24.12
Central	24.26	22.82	22.82	24.26
West	28.41	28.35	28.35	28.41
Parents' Education				
Less than high school	4.85	3.93	3.93	4.85
High school	16.03	15.66	15.66	16.03
Greater than high school	7.38	8.01	8.01	7.38
Graduated college	40.02	42.91	42.91	40.02
Unknown	31.63	29.39	29.39	31.63
Size and Type of Community				
Rural	7.50	7.06	7.06	7.50
Disadvantaged urban	8.77	8.84	8.84	8.77
Advantaged urban	11.83	13.74	13.74	11.83
Big city	7.40	7.84	7.84	7.40
Fringe	8.57	8.32	8.32	8.57
Medium city	14.75	14.17	14.17	14.75
Small place	41.18	40.04	40.04	41.18
Estimated Total Population	3637675	2358141	2358141	3637675

Table 16-45

Weighted Percentage of Students in Bridge to 1986 Sample (Age Only)
by Type of Eligibility and Subgroup Classification
Age 13

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Sex				
Male	49.83	45.94	45.94	49.83
Female	50.17	54.06	54.06	50.17
Race/Ethnicity				
White	70.08	72.62	72.62	70.08
Black	14.77	13.42	13.42	14.77
Hispanic	10.71	8.82	8.82	10.71
Other	4.44	5.15	5.15	4.44
Region				
Northeast	24.30	25.89	25.89	24.30
Southeast	23.10	23.53	23.53	23.10
Central	23.50	22.08	22.08	23.50
West	29.10	28.50	28.50	29.10
Parents' Education				
Less than high school	7.59	6.19	6.19	7.59
High school	26.75	24.85	24.85	26.75
Greater than high school	16.77	19.29	19.29	16.77
Graduated college	40.73	43.87	43.87	40.73
Unknown	7.93	5.63	5.63	7.93
Size and Type of Community				
Rural	9.75	8.74	8.74	9.75
Disadvantaged urban	11.03	9.64	9.64	11.03
Advantaged urban	9.68	10.63	10.63	9.68
Big city	7.61	8.81	8.81	7.61
Fringe	9.63	10.36	10.36	9.63
Medium city	11.28	11.38	11.38	11.28
Small place	41.02	40.45	40.45	41.02
Estimated Total Population	3105980	1967661	1967661	3105980

Table 16-46

Weighted Percentage of Students in Bridge to 1986 Sample (Age Only)
by Type of Eligibility and Subgroup Classification
Age 17

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Sex				
Male	48.64	46.26	46.26	48.64
Female	51.36	53.74	53.74	51.36
Race/Ethnicity				
White	71.37	74.76	74.76	71.37
Black	15.32	13.63	13.63	15.32
Hispanic	9.16	7.95	7.95	9.16
Other	4.16	3.67	3.67	4.16
Region				
Northeast	22.15	20.86	20.86	22.15
Southeast	24.33	23.10	23.10	24.33
Central	25.87	28.16	28.16	25.87
West	27.66	27.88	27.88	27.66
Parents' Education				
Less than high school	7.89	6.10	6.10	7.89
High school	26.36	24.53	24.53	26.36
Greater than high school	23.72	26.00	26.00	23.72
Graduated college	38.88	41.21	41.21	38.88
Unknown	2.97	2.09	2.09	2.97
Size and Type of Community				
Rural	11.54	10.84	10.84	11.54
Disadvantaged urban	9.35	7.42	7.42	9.35
Advantaged urban	10.48	11.19	11.19	10.48
Big city	8.02	7.86	7.86	8.02
Fringe	11.57	11.81	11.81	11.57
Medium city	13.24	13.26	13.26	13.24
Small place	35.81	37.63	37.63	35.81
Estimated Total Population	3002755	2108602	2108602	3002755

Table 16-47

Weighted Percentage of Students in Bridge to 1986 Sample (Age and Grade) by Type of Eligibility and Subgroup Classification
Age 17/Grade 11

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Sex				
Male	49.67	49.43	46.75	50.87
Female	50.33	50.57	53.25	49.13
Race/Ethnicity				
White	71.29	71.35	75.59	69.29
Black	15.42	15.43	13.00	16.58
Hispanic	9.12	8.75	7.42	9.66
Other	4.17	4.46	3.99	4.48
Region				
Northeast	22.59	22.13	22.49	22.29
Southeast	23.62	23.84	22.04	24.54
Central	25.39	25.77	28.05	24.41
West	28.40	28.26	27.43	28.76
Parents' Education				
Less than high school	9.34	8.43	7.24	9.66
High school	25.77	25.14	24.28	26.01
Greater than high school	24.22	24.44	25.50	23.77
Graduated college	37.34	38.48	40.62	36.63
Unknown	3.00	3.24	2.18	3.57
Size and Type of Community				
Rural	12.92	12.60	12.14	13.06
Disadvantaged urban	8.98	8.79	6.99	9.78
Advantaged urban	10.86	11.46	11.82	10.85
Big city	8.20	8.48	7.76	8.62
Fringe	10.98	10.93	11.17	10.85
Medium city	13.61	13.91	13.97	13.66
Small place	34.45	33.82	36.15	33.17
Estimated Total Population	2989051	3040093	1945083	4084061

Table 16-48

Weighted Percentage of Students in Long-term Trend Sample
by Type of Eligibility and Subgroup Classification
Age 9

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Sex				
Male	49.78	46.04	46.04	49.78
Female	50.22	53.96	53.96	50.22
Race/Ethnicity				
White	69.81	70.21	70.21	69.81
Black	15.33	15.33	15.33	15.33
Hispanic	10.67	10.33	10.33	10.67
Other	4.18	4.13	4.13	4.18
Region				
Northeast	23.41	25.45	25.45	23.41
Southeast	24.15	23.05	23.05	24.15
Central	24.46	23.38	23.38	24.46
West	27.99	28.12	28.12	27.99
Parents' Education				
Less than high school	4.67	3.98	3.98	4.67
High school	15.39	15.63	15.63	15.39
Greater than high school	7.46	8.27	8.27	7.46
Graduated college	39.46	41.67	41.67	39.46
Unknown	32.87	30.36	30.36	32.87
Size and Type of Community				
Rural	9.56	9.41	9.41	9.56
Disadvantaged urban	9.11	8.90	8.90	9.11
Advantaged urban	10.36	12.38	12.38	10.36
Big city	7.75	8.42	8.42	7.75
Fringe	9.18	9.31	9.31	9.18
Medium city	16.14	14.29	14.29	16.14
Small place	37.89	37.30	37.30	37.89
Estimated Total Population	3599721	2299759	2299759	3599721

Table 16-49

Weighted Percentage of Students in Long-term Trend Sample
by Type of Eligibility and Subgroup Classification
Age 13

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Sex				
Male	49.13	44.39	44.39	49.13
Female	50.87	55.61	55.61	50.87
Race/Ethnicity				
White	69.98	74.24	74.24	69.98
Black	14.98	11.54	11.54	14.98
Hispanic	10.73	9.58	9.58	10.73
Other	4.31	4.64	4.64	4.31
Region				
Northeast	23.38	26.66	26.66	23.38
Southeast	25.30	22.98	22.98	25.30
Central	22.48	21.52	21.52	22.48
West	28.83	28.84	28.84	28.83
Parents' Education				
Less than high school	7.51	6.33	6.33	7.51
High school	27.18	25.17	25.17	27.18
Greater than high school	17.56	19.39	19.39	17.56
Graduated college	39.38	42.48	42.48	39.38
Unknown	8.02	6.25	6.25	8.02
Size and Type of Community				
Rural	10.48	9.27	9.27	10.48
Disadvantaged urban	10.05	9.23	9.23	10.05
Advantaged urban	11.68	12.55	12.55	11.68
Big city	7.62	9.09	9.09	7.62
Fringe	8.90	10.16	10.16	8.90
Medium city	10.68	10.57	10.57	10.68
Small place	40.59	39.13	39.13	40.59
Estimated Total Population	3116104	1904527	1904527	3116104

Table 16-50

Weighted Percentage of Students in Long-term Trend Sample by Type of Eligibility and Subgroup Classification
Age 17

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Sex				
Male	48.65	45.62	45.62	48.65
Female	51.35	54.38	54.38	51.35
Race/Ethnicity				
White	71.38	76.15	76.15	71.38
Black	15.39	13.21	13.21	15.39
Hispanic	9.09	7.11	7.11	9.09
Other	4.14	3.53	3.53	4.14
Region				
Northeast	22.27	21.51	21.51	22.27
Southeast	23.29	22.75	22.75	23.29
Central	25.85	27.60	27.60	25.85
West	28.59	28.13	28.13	28.59
Parents' Education				
Less than high school	8.49	6.62	6.62	8.49
High school	25.18	22.48	22.48	25.18
Greater than high school	24.66	26.08	26.08	24.66
Graduated college	39.55	43.10	43.10	39.55
Unknown	1.54	1.08	1.08	1.54
Size and Type of Community				
Rural	12.59	12.20	12.20	12.59
Disadvantaged urban	9.15	7.39	7.39	9.15
Advantaged urban	10.10	10.66	10.66	10.10
Big city	8.05	7.76	7.76	8.05
Fringe	11.26	11.90	11.90	11.26
Medium city	13.50	13.47	13.47	13.50
Small place	35.34	36.62	36.62	35.34
Estimated Total Population	3000558	2120949	2120949	3000558

Table 16-51

Weighted Proportion of Excluded Students in Main Sample
by Type of Eligibility and Subgroup Classification*
Age 9/Grade 4

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Sex				
Male	65.08	66.24	63.05	66.06
Female	34.92	33.76	36.95	33.94
Race/Ethnicity				
White	52.04	52.69	46.40	53.37
Black	18.93	16.78	13.89	18.64
Hispanic	15.43	18.03	25.03	15.17
Other	13.60	12.50	14.68	12.82
Region				
Northeast	24.81	22.40	27.95	22.95
Southeast	18.86	19.03	8.97	20.68
Central	22.75	22.23	14.81	23.84
West	33.58	36.33	48.27	32.52
Size and Type of Community				
Rural	9.98	7.29	5.61	9.28
Disadvantaged urban	21.72	18.84	26.12	19.39
Advantaged urban	6.32	4.79	7.91	5.21
Big city	6.40	7.60	9.11	6.58
Fringe	14.27	17.43	18.37	15.27
Medium city	12.36	13.96	10.14	13.62
Small place	28.95	30.09	22.75	30.65
Estimated Total Population	410238	353248	113156	650330

* Data on parents' education were not collected for excluded students.

Table 16-52

Weighted Proportion of Excluded Students in Main Sample
by Type of Eligibility and Subgroup Classification*
Age 13/Grade 8

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Sex				
Male	63.48	65.49	57.85	65.39
Female	36.52	34.51	42.15	34.61
Race/Ethnicity				
White	47.96	50.23	45.83	49.50
Black	18.43	19.08	15.38	19.23
Hispanic	18.62	19.64	22.95	18.55
Other	14.99	11.06	15.85	12.72
Region				
Northeast	17.40	16.79	22.41	16.34
Southeast	19.77	16.12	7.80	19.52
Central	27.73	31.95	29.48	29.77
West	35.10	35.15	40.31	34.37
Size and Type of Community				
Rural	10.72	8.98	6.47	10.39
Disadvantaged urban	15.89	16.37	18.98	15.71
Advantaged urban	5.13	3.70	4.22	4.48
Big city	12.29	12.72	18.21	11.67
Fringe	8.57	8.16	9.87	8.16
Medium city	18.58	18.05	17.82	18.40
Small place	28.82	32.02	24.43	31.20
Estimated Total Population	352527	320265	85478	587313

* Data on parents' education were not collected for excluded students.

Table 16-53

Weighted Proportion of Excluded Students in Main Sample
by Type of Eligibility and Subgroup Classification*
Age 17/Grade 12

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Sex				
Male	65.49	62.15	50.60	65.79
Female	34.51	37.85	49.40	34.21
Race/Ethnicity				
White	48.17	56.92	54.20	52.05
Black	22.17	19.38	12.69	22.00
Hispanic	15.81	10.83	14.83	13.26
Other	13.85	12.87	18.28	12.69
Region				
Northeast	15.47	15.10	14.93	15.35
Southeast	24.87	21.13	22.03	23.25
Central	22.32	30.28	29.47	25.62
West	37.34	33.49	33.56	35.79
Size and Type of Community				
Rural	10.66	14.73	11.26	12.78
Disadvantaged urban	21.81	15.90	17.63	19.21
Advantaged urban	4.39	3.57	2.23	4.25
Big city	5.32	5.36	6.21	5.21
Fringe	10.36	9.52	13.51	9.46
Medium city	16.79	17.47	14.81	17.44
Small place	30.66	33.45	34.35	31.65
Estimated Total Population	206859	186490	48873	344476

* Data on parents' education were not collected for excluded students.

Table 16-54

Weighted Proportion of Excluded Students in Bridge Samples
by Type of Eligibility and Subgroup Classification*
Age 9/Grade 4

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Sex				
Male	63.19	68.40	58.53	67.44
Female	36.81	31.60	41.47	32.56
Race/Ethnicity				
White	53.44	50.11	43.07	53.65
Black	13.63	16.09	10.52	15.83
Hispanic	17.65	19.35	29.05	16.20
Other	6.98	5.07	9.54	5.24
Region				
Northeast	23.67	21.58	25.21	22.04
Southeast	16.74	26.22	12.31	23.58
Central	23.23	21.43	17.13	23.46
West	36.36	30.78	45.35	30.93
Size and Type of Community				
Rural	4.20	8.05	4.54	6.51
Disadvantaged urban	17.56	15.46	19.25	15.89
Advantaged urban	9.96	7.19	9.93	8.25
Big city	9.12	8.69	17.98	6.91
Fringe	9.75	8.97	11.00	8.99
Medium city	12.41	12.11	10.27	12.70
Small place	37.00	39.53	27.04	40.75
Estimated Total Population	203416	209561	74332	338645

* Data on parents' education were not collected for excluded students.

Table 16-55

Weighted Proportion of Excluded Students in Bridge Samples
by Type of Eligibility and Subgroup Classification*
Age 13/Grade 8

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Sex				
Male	62.83	62.49	58.93	63.26
Female	37.17	37.51	41.07	36.74
Race/Ethnicity				
White	48.04	50.30	48.61	49.26
Black	19.75	18.00	13.87	19.69
Hispanic	19.91	22.75	26.25	20.53
Other	12.30	8.96	11.27	10.52
Region				
Northeast	21.92	18.20	20.97	19.91
Southeast	29.19	26.32	31.38	27.16
Central	16.64	19.34	8.70	19.51
West	32.26	36.14	38.95	33.42
Size and Type of Community				
Rural	10.60	11.48	16.31	10.18
Disadvantaged urban	20.86	16.63	25.26	17.68
Advantaged urban	9.08	5.13	6.99	7.13
Big city	8.93	10.27	13.56	8.95
Fringe	6.87	8.84	6.61	8.06
Medium city	10.30	13.54	5.29	13.00
Small place	33.37	34.12	25.99	35.01
Estimated Total Population	173041	173271	48595	297717

* Data on parents' education were not collected for excluded students.

Table 16-56

Weighted Proportion of Excluded Students in Bridge Samples
by Type of Eligibility and Subgroup Classification*
Age 17/Grade 11

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Sex				
Male	65.29	67.81	61.11	67.24
Female	34.71	32.19	38.89	32.76
Race/Ethnicity				
White	50.67	48.01	53.59	48.82
Black	13.48	20.95	13.75	17.53
Hispanic	21.25	12.05	12.96	17.37
Other	14.59	18.98	19.70	16.28
Region				
Northeast	14.71	17.17	13.26	16.25
Southeast	22.13	23.39	24.14	22.54
Central	20.06	22.24	25.18	20.55
West	43.10	37.21	37.42	40.66
Size and Type of Community				
Rural	10.77	8.43	12.11	9.31
Disadvantaged urban	16.21	11.10	9.71	14.32
Advantaged urban	8.52	8.14	5.94	8.67
Big city	9.44	12.07	9.03	10.93
Fringe	10.07	11.84	11.75	10.80
Medium city	12.63	14.40	10.90	13.84
Small place	32.36	34.03	40.56	32.14
Estimated Total Population	140448	129609	32741	237316

* Data on parents' education were not collected for excluded students.

Table 16-57

Number of Students in NAEP Main Sample with Proficiency Scores by Type of Eligibility Age 9/Grade 4

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Reading	6074	6323	3917	8480
Reading Answer Booklet Bridge	2618	2718	1721	3615
Mathematics	6357	6467	4034	8790
Science	5953	6314	3849	8418
TOTAL	21002	21822	13521	29303

Table 16-58

Number of Students in NAEP Main Sample
with Proficiency Scores by Type of Eligibility
Age 13/Grade 8

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Reading	6278	6510	4063	8725
Mathematics	6201	6473	4040	8634
Science	6217	6531	4039	8709
TOTAL	18696	19514	12142	26068

Table 16-59

Number of Students in NAEP Main Sample
with Proficiency Scores by Type of Eligibility
Age 17/Grade 12

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Reading	6328	6258	4235	8351
Mathematics	6410	6311	4315	8406
Science	6481	6337	4373	8445
TOTAL	19219	18906	12923	25202

Table 16-60

Number of Students in Bridge to 1984 Sample with Proficiency Scores by Type of Eligibility

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Reading/Writing				
Age 9/Grade 4	4268	4367	2709	5926
Age 13/Grade 8	4609	4665	3041	6233
Age 17/Grade 11	4383	4385	3154	5614
TOTAL	13260	13417	8904	17773

Table 16-61

Number of Students in Bridge to 1986 Sample (Age Only) with Proficiency Scores by Type of Eligibility

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
Mathematics/Science				
Age 9	6235	4016	4016	6235
Age 13	6649	4330	4330	6649
Age 17	4411	3173	3173	4411
TOTAL	17295	11519	11519	17295

Table 16-62

Weighted Proficiency Means, Standard Deviations, and Percentiles (with Standard Errors) for Main Focused-BIB Reading Samples, by Subgroup, Grade 4

	MEAN	STD DEV	5TH	10TH	25TH	50TH	75TH	90TH	95TH
-- TOTAL --	233.0(0.9)	42.0(0.4)	160.8(1.0)	177.3(1.9)	205.5(1.1)	235.1(0.9)	261.9(1.0)	285.5(1.1)	299.0(0.9)
GENDER									
MALE	228.1(1.1)	43.5(0.6)	154.6(2.0)	169.7(1.5)	198.3(0.9)	230.2(1.4)	258.6(1.1)	283.1(1.6)	297.0(1.4)
FEMALE	238.0(1.0)	39.9(0.7)	169.8(2.0)	186.0(1.7)	212.5(1.1)	239.3(1.0)	264.8(1.0)	287.3(2.1)	300.9(2.1)
RACE/ETHNICITY									
WHITE	241.2(0.9)	39.8(0.5)	172.5(2.8)	188.6(1.4)	215.6(1.4)	242.8(0.9)	268.3(0.9)	290.4(0.8)	304.0(1.6)
BLACK	209.4(1.9)	39.7(1.0)	142.0(4.3)	157.2(2.2)	182.7(3.9)	211.3(2.1)	236.5(1.6)	258.9(2.6)	273.1(5.2)
HISPANIC	212.9(1.6)	40.6(1.2)	144.7(2.2)	159.6(3.0)	185.8(2.1)	214.7(1.8)	241.3(1.6)	263.8(1.8)	276.8(2.4)
ASIAN/PACIFIC AMERICAN	236.6(3.0)	41.8(1.8)	165.3(8.4)	179.6(3.7)	208.1(3.1)	239.4(6.0)	265.4(4.1)	288.8(4.9)	301.5(4.6)
AMER IND/ALASKAN NATV	229.1(2.5)	37.7(2.1)	165.7(5.1)	177.9(5.0)	205.6(7.5)	230.8(5.7)	254.0(2.5)	277.0(11.2)	290.0(7.8)
PARENTS' EDUCATION LEVEL									
LESS THAN H.S.	214.4(1.9)	40.0(1.6)	146.9(5.9)	163.3(2.3)	186.8(2.0)	215.9(3.7)	242.3(2.4)	264.9(4.3)	278.7(2.7)
GRADUATED H.S.	227.4(1.4)	41.3(1.2)	155.0(4.8)	170.8(3.6)	201.1(2.2)	229.1(2.1)	255.4(1.3)	280.3(3.1)	292.9(2.6)
SOME EDUC AFTER H.S.	241.6(2.0)	41.8(1.3)	168.2(9.5)	183.7(3.0)	214.1(1.5)	245.8(3.1)	270.0(3.5)	291.9(3.9)	305.3(6.1)
GRADUATED COLLEGE	242.1(1.3)	41.2(0.6)	171.2(2.8)	188.0(1.8)	215.1(1.6)	243.3(1.3)	270.7(1.7)	293.4(1.5)	307.4(2.4)
UNKNOWN	226.9(1.0)	40.9(0.9)	155.6(3.4)	173.1(2.0)	200.4(1.3)	229.2(1.5)	255.3(1.2)	277.2(1.7)	290.1(1.2)

416

417

Table 16-63

Weighted Proficiency Means, Standard Deviations, and Percentiles (with Standard Errors) for Main Focused-BIB Reading Samples, by Subgroup, Grade 8

	MEAN	STD DEV	5TH	10TH	25TH	50TH	75TH	90TH	95TH
-- TOTAL --	260.6(1.0)	37.4(0.6)	196.0(1.6)	211.5(1.6)	236.5(1.4)	261.9(1.4)	286.2(0.8)	306.5(1.1)	319.5(2.0)
GENDER									
MALE	252.7(1.2)	38.2(0.6)	187.7(2.4)	201.8(1.6)	227.5(2.3)	253.5(1.2)	279.4(1.3)	301.1(1.3)	313.8(2.2)
FEMALE	268.3(1.0)	34.8(0.9)	209.7(2.4)	224.0(1.9)	245.5(1.7)	269.7(1.1)	291.1(1.0)	311.4(1.2)	324.2(2.4)
RACE/ETHNICITY									
WHITE	265.8(1.1)	37.0(0.7)	200.3(1.3)	216.6(1.9)	242.3(1.4)	268.3(1.4)	290.9(1.2)	310.8(1.4)	323.5(1.3)
BLACK	246.4(1.6)	33.5(1.2)	188.7(3.1)	204.4(3.3)	226.2(3.3)	246.5(1.2)	268.6(3.2)	288.5(1.5)	300.6(2.6)
HISPANIC	243.3(1.8)	34.8(1.1)	180.9(4.0)	196.1(4.4)	221.0(2.0)	244.9(2.1)	267.2(1.8)	286.6(2.6)	297.2(5.4)
ASIAN/PACIFIC AMERICAN	277.3(3.4)	35.2(1.8)	218.4(11.4)	232.6(11.9)	253.2(6.9)	277.8(3.6)	302.7(6.9)	321.0(3.6)	331.5(5.7)
AMER IND/ALASKAN NATV	248.2(3.3)	31.6(3.1)	192.0(9.7)	205.4(19.8)	229.6(4.7)	248.3(3.8)	268.2(3.5)	290.6(10.5)	300.5(6.7)
PARENTS' EDUCATION LEVEL									
LESS THAN H.S.	245.1(2.0)	34.0(1.3)	185.7(3.7)	198.5(4.1)	223.7(3.0)	246.3(5.0)	269.8(1.7)	287.3(3.0)	297.2(2.7)
GRADUATED H.S.	250.9(1.3)	34.3(0.8)	191.1(3.1)	205.8(1.8)	229.6(2.2)	252.5(1.2)	274.3(0.6)	292.7(1.2)	304.4(1.6)
SOME EDUC AFTER H.S.	265.1(1.4)	35.8(1.1)	200.4(3.1)	218.5(1.3)	243.0(2.5)	268.2(1.9)	290.2(1.3)	307.6(1.1)	318.2(4.6)
GRADUATED COLLEGE	271.8(1.2)	36.8(0.8)	208.0(3.3)	222.9(2.9)	248.1(1.3)	273.4(1.4)	296.5(1.5)	317.3(1.9)	329.8(2.0)
UNKNOWN	242.4(2.3)	35.4(1.8)	182.5(3.7)	195.5(2.9)	218.7(3.9)	241.8(2.0)	266.6(4.5)	289.9(6.4)	301.4(3.5)

Table 16-64

Weighted Proficiency Means, Standard Deviations, and Percentiles (with Standard Errors) for Main Focused-BIB Reading Samples, by Subgroup, Grade 12

	MEAN	STD DEV	5TH	10TH	25TH	50TH	75TH	90TH	95TH
-- TOTAL --	288.5(1.0)	35.0(0.6)	227.4(1.7)	243.1(1.4)	266.4(1.1)	290.2(0.7)	312.5(1.2)	331.6(1.1)	342.9(2.0)
GENDER									
MALE	283.3(1.3)	37.1(0.8)	217.8(2.3)	234.5(2.4)	259.7(3.2)	285.5(1.8)	309.6(1.2)	329.3(1.1)	340.1(1.6)
FEMALE	293.4(0.9)	32.2(0.6)	238.8(2.9)	251.8(0.9)	273.0(1.3)	294.1(1.1)	315.2(0.7)	333.5(1.4)	345.1(3.1)
RACE/ETHNICITY									
WHITE	294.0(1.0)	33.6(0.6)	235.2(3.0)	250.8(1.7)	273.5(1.3)	295.8(1.3)	317.0(1.0)	334.9(1.5)	346.3(1.9)
BLACK	269.2(1.9)	31.1(1.1)	214.9(8.0)	229.4(4.5)	249.4(3.5)	270.0(2.2)	290.4(2.7)	308.1(2.5)	318.6(3.1)
HISPANIC	271.4(2.5)	36.9(1.5)	208.7(6.7)	222.2(3.3)	247.5(3.5)	273.8(4.2)	297.2(4.0)	317.2(3.7)	328.7(3.2)
ASIAN/PACIFIC AMERICAN	290.6(3.9)	33.8(2.1)	235.5(4.7)	246.0(5.3)	269.0(2.5)	289.7(7.2)	313.1(5.4)	336.7(3.4)	345.7(11.7)
AMER IND/ALASKAN NATV	283.4(7.6)	33.8(5.5)	226.3(24.6)	239.4(17.8)	263.7(4.8)	282.9(12.8)	308.6(11.4)	325.5(10.0)	338.3(10.9)
PARENTS' EDUCATION LEVEL									
LESS THAN H.S.	269.6(1.9)	31.5(1.6)	215.6(3.1)	229.4(8.1)	250.2(4.0)	271.1(3.2)	291.4(2.1)	309.0(2.4)	317.6(4.8)
GRADUATED H.S.	279.2(1.0)	32.2(0.9)	224.3(3.4)	238.9(2.4)	259.6(2.6)	280.6(1.1)	301.0(1.2)	319.4(1.0)	329.3(1.6)
SOME EDUC AFTER H.S.	291.8(1.1)	31.5(0.8)	236.0(4.1)	251.0(1.5)	273.0(2.4)	292.9(1.5)	312.6(1.0)	330.1(1.6)	341.1(2.0)
GRADUATED COLLEGE	297.3(1.1)	35.6(0.7)	232.7(2.8)	249.3(3.1)	276.0(1.7)	300.2(1.1)	322.0(1.2)	339.9(1.3)	350.6(1.7)
UNKNOWN	256.9(4.4)	32.6(2.4)	197.8(9.4)	212.2(14.4)	237.9(6.3)	256.9(7.[illegible])	279.1(9.6)	298.5(9.0)	312.6(17.8)

420

421

Table 16-65

Weighted Proficiency Means, Standard Deviations, and Percentiles (with Standard Errors) for Bridge to 1984 Reading Samples, by Subgroup, Age 9

	MEAN	STD DEV	5TH	10TH	25TH	50TH	75TH	90TH	95TH
-- TOTAL --	209.2(1.2)	44.7(0.8)	134.8(3.2)	150.1(1.9)	178.7(1.8)	210.3(1.5)	240.3(1.8)	265.7(1.8)	280.4(1.3)
GENDER									
MALE	204.0(1.7)	45.1(1.0)	129.6(5.8)	145.1(1.9)	172.2(2.8)	204.4(2.2)	236.1(1.9)	261.7(2.6)	276.1(5.6)
FEMALE	214.5(1.2)	43.6(1.3)	140.6(3.9)	156.8(3.2)	185.7(1.7)	215.9(1.3)	244.4(1.9)	269.4(1.9)	284.1(2.1)
OBSERVED RACE/ETHNICITY									
WHITE	217.0(1.3)	42.9(1.0)	144.2(3.2)	160.0(1.5)	188.0(2.8)	218.4(2.1)	246.7(2.3)	270.9(2.1)	285.3(2.6)
BLACK	181.8(2.9)	41.7(1.7)	115.0(4.7)	128.9(3.9)	152.5(3.2)	181.8(3.1)	210.5(2.4)	236.3(2.7)	250.7(6.9)
HISPANIC	189.4(2.3)	39.7(1.6)	125.4(8.9)	139.0(4.3)	160.8(1.9)	189.3(3.5)	218.9(4.0)	239.3(5.7)	253.2(6.7)
ASIAN/PACIFIC AMERICAN	207.0(4.6)	40.5(3.9)	136.7(11.6)	150.1(11.5)	179.2(6.9)	210.5(3.9)	235.6(12.2)	256.5(7.3)	271.7(6.0)
AMER IND/ALASKAN NATV	182.8(16.6)	45.3(16.2)	131.1(34.9)	131.8(35.0)	134.3(39.2)	178.7(42.7)	209.0(46.5)	251.0(9.8)	250.0(25.8)
PARENTS' EDUCATION LEVEL									
LESS THAN H.S.	192.6(3.2)	42.7(2.5)	124.1(14.9)	135.4(7.8)	162.7(4.6)	192.8(6.4)	222.2(5.5)	248.9(3.9)	264.1(7.9)
GRADUATED H.S.	209.1(1.8)	42.6(1.5)	138.4(3.5)	151.7(3.2)	179.9(4.4)	210.2(3.0)	238.9(2.9)	263.0(7.0)	278.1(4.6)
SOME EDUC AFTER H.S.	210.4(3.4)	48.1(3.1)	130.0(9.8)	146.2(9.7)	175.3(6.2)	212.6(7.1)	243.4(7.9)	270.8(5.5)	284.7(9.0)
GRADUATED COLLEGE	218.8(2.0)	44.2(1.2)	143.7(2.5)	159.0(4.1)	188.3(3.3)	221.0(2.2)	249.6(3.1)	273.9(2.7)	289.2(2.5)
UNKNOWN	201.4(1.5)	43.5(1.2)	128.5(3.5)	143.8(3.3)	172.6(2.0)	202.7(1.7)	230.9(1.5)	256.3(2.6)	270.5(5.1)

422

423

Table 16-66

Weighted Proficiency Means, Standard Deviations, and Percentiles (with Standar(i rrors)
for Bridge to 1984 Reading Samples, by Subgroup, Age 13

	MEAN	STD DEV	5TH	10TH	25TH	50TH	75TH	90TH	95TH
-- TOTAL --	256.8(0.8)	36.0(0.6)	195.7(1.9)	209.8(1.8)	233.2(1.0)	257.3(0.9)	281.5(0.8)	302.0(1.0)	314.4(1.3)
GENDER									
MALE	250.5(1.1)	36.0(0.7)	189.7(2.2)	202.8(1.4)	226.9(1.9)	251.9(1.3)	275.3(1.2)	295.3(1.2)	307.4(3.2)
FEMALE	263.1(1.1)	34.8(0.7)	205.3(3.1)	217.9(2.0)	240.0(1.9)	263.0(1.6)	286.6(1.1)	308.1(1.5)	319.4(2.5)
OBSERVED RACE/ETHNICITY									
WHITE	262.3(0.9)	34.5(0.6)	204.1(2.2)	217.3(1.7)	239.6(1.7)	262.6(1.4)	285.6(1.2)	306.0(2.4)	318.1(2.7)
BLACK	241.5(2.2)	35.3(1.5)	182.3(5.3)	194.3(7.3)	217.0(3.2)	242.5(4.0)	265.7(2.5)	285.9(4.9)	298.9(3.0)
HISPANIC	237.8(2.3)	35.9(1.3)	178.0(9.6)	191.3(4.9)	214.1(4.1)	238.6(4.1)	262.2(3.1)	283.8(6.0)	295.9(4.5)
ASIAN/PACIFIC AMERICAN	253.5(5.7)	33.7(3.2)	197.1(9.2)	211.1(4.9)	231.7(9.7)	252.6(4.9)	278.4(10.5)	296.6(4.5)	308.6(5.6)
AMER IND/ALASKAN NATV	235.5(15.7)	30.9(10.4)	204.5(18.2)	204.5(18.2)	204.6(18.3)	234.4(30.9)	258.3(52.8)	289.6(93.8)	291.8(26.6)
PARENTS' EDUCATION LEVEL									
LESS THAN H.S.	240.8(1.8)	33.8(2.0)	183.0(3.0)	195.9(4.6)	218.9(4.5)	241.2(1.4)	262.5(3.1)	283.2(5.2)	296.4(5.5)
GRADUATED H.S.	251.4(0.9)	33.1(0.9)	195.3(2.7)	208.2(2.9)	229.2(1.3)	252.1(1.1)	274.0(2.5)	294.0(1.5)	305.8(2.1)
SOME EDUC AFTER H.S.	267.4(1.7)	33.3(1.6)	208.7(9.3)	223.9(3.5)	247.2(1.8)	268.5(1.3)	289.2(2.3)	309.9(6.1)	321.4(5.9)
GRADUATED COLLEGE	266.8(1.1)	35.1(0.9)	206.0(3.0)	221.2(2.6)	244.6(1.8)	267.3(1.3)	290.5(1.5)	310.1(2.2)	323.0(2.9)
UNKNOWN	237.7(1.9)	35.1(1.7)	180.5(9.6)	192.5(4.9)	214.5(3.2)	237.4(2.2)	261.5(1.8)	283.3(2.7)	294.5(2.2)

424 425

Table 16-67

Weighted Proficiency Means, Standard Deviations, and Percentiles (with Standard Errors)
for Bridge to 1984 Reading Samples, by Subgroup, Age 17

	MEAN	STD DEV	5TH	10TH	25TH	50TH	75TH	90TH	95TH
-- TOTAL --	290.2(1.1)	41.3(0.7)	220.0(2.3)	236.9(3.1)	263.5(1.3)	291.1(1.3)	318.6(1.5)	342.7(2.1)	356.0(1.7)
GENDER									
MALE	284.0(.6)	42.6(0.8)	209.4(3.2)	228.2(3.4)	257.3(1.9)	285.9(2.1)	313.2(2.1)	338.4(2.3)	351.9(1.6)
FEMALE	296.5(1.2)	38.8(0.8)	232.3(3.8)	247.0(2.1)	270.5(2.3)	296.6(1.2)	323.5(1.5)	346.3(2.5)	359.4(2.7)
OBSERVED RACE/ETHNICITY									
WHITE	296.6(1.2)	39.6(0.6)	228.5(2.5)	246.2(2.5)	271.1(1.4)	297.5(1.2)	323.8(1.9)	347.1(1.6)	359.7(1.7)
BLACK	267.3(2.3)	39.2(2.2)	201.3(7.9)	217.4(4.0)	242.4(3.9)	268.4(1.9)	293.7(2.7)	316.2(4.8)	330.5(11.0)
HISPANIC	274.8(3.6)	40.7(2.7)	205.9(11.1)	224.3(12.0)	250.4(8.3)	276.3(3.2)	302.6(4.9)	326.5(3.2)	339.4(11.2)
ASIAN/PACIFIC AMERICAN	290.7(3.3)	42.3(4.4)	219.8(18.6)	236.6(11.0)	262.4(5.5)	293.1(4.2)	322.2(13.5)	342.3(5.6)	357.6(25.2)
AMER IND/ALASKAN NATV	267.7(9.3)	24.7(6.2)	221.6(11.2)	238.3(5.4)	252.4(18.3)	263.6(24.7)	281.6(30.7)	300.4(30.4)	309.3(21.8)
PARENTS' EDUCATION LEVEL									
LESS THAN H.S.	269.7(2.8)	37.8(1.7)	207.0(26.3)	220.5(10.5)	244.8(3.1)	271.1(3.9)	295.2(6.7)	316.7(3.3)	330.0(3.9)
GRADUATED H.S.	282.9(1.4)	38.5(1.0)	217.6(4.2)	233.4(3.7)	258.5(1.6)	282.9(1.4)	308.8(2.8)	332.7(3.1)	345.9(3.8)
SOME EDUC AFTER H.S.	295.0(1.9)	38.0(1.9)	230.4(3.8)	246.0(1.7)	270.3(4.5)	296.3(2.4)	321.3(3.4)	342.3(2.3)	355.5(4.5)
GRADUATED COLLEGE	302.2(1.5)	39.4(1.1)	235.0(4.3)	251.5(3.5)	276.1(2.4)	303.6(1.4)	330.1(2.0)	352.2(1.2)	364.5(2.3)
UNKNOWN	245.9(5.7)	44.2(4.6)	167.9(18.1)	186.0(21.5)	218.6(6.3)	247.3(5.3)	276.9(3.9)	302.5(8.3)	314.9(8.3)

426

427

Table 16-68

Weighted Proficiency Means, Standard Deviations, and Percentiles (with Standard Errors)
for Main Focused-BIB Mathematics Samples, by Subgroup, Grade 4

	MEAN	STD DEV	5TH	10TH	25TH	50TH	75TH	90TH	95TH
TOTAL --	215.8(0.7)	28.1(0.3)	168.4(1.1)	179.0(1.4)	197.1(1.2)	216.8(0.8)	235.1(0.6)	251.4(0.9)	260.8(0.9)
GENDER									
MALE	216.7(0.8)	28.4(0.5)	167.9(1.6)	179.1(1.6)	197.8(1.1)	218.0(1.2)	236.2(0.9)	252.3(1.3)	261.3(1.4)
FEMALE	214.9(0.8)	27.6(0.5)	168.7(1.7)	179.0(1.4)	196.6(1.1)	215.3(0.8)	233.9(1.4)	250.3(1.1)	260.3(1.3)
RACE/ETHNICITY									
WHITE	222.7(0.7)	25.4(0.4)	180.2(1.9)	189.9(1.4)	205.7(1.3)	223.0(0.9)	239.9(0.9)	255.1(1.0)	263.7(1.6)
BLACK	194.1(1.3)	25.3(0.7)	152.1(2.1)	162.6(3.5)	177.1(1.7)	194.5(1.3)	211.0(1.0)	226.2(1.4)	235.4(1.8)
HISPANIC	200.5(1.4)	27.4(0.9)	154.4(3.7)	165.5(1.6)	182.2(2.2)	201.2(1.5)	219.2(1.9)	236.0(1.7)	245.4(3.2)
ASIAN/PACIFIC AMERICAN	228.2(2.8)	27.4(1.6)	180.8(4.7)	191.2(5.1)	210.0(7.3)	229.3(6.2)	247.8(4.0)	263.2(3.7)	269.5(2.4)
AMER IND/ALASKAN NATV	210.5(2.5)	23.8(1.5)	171.1(4.7)	179.4(3.6)	193.2(2.5)	211.0(4.1)	228.8(2.4)	239.6(1.5)	245.8(7.7)
PARENTS' EDUCATION LEVEL									
LESS THAN H.S.	205.2(2.0)	25.6(1.1)	162.0(4.5)	172.8(3.3)	187.7(3.9)	205.7(2.1)	223.0(2.1)	237.3(3.2)	244.8(2.6)
GRADUATED H.S.	211.3(1.2)	25.4(0.8)	167.8(1.9)	178.8(2.2)	194.8(1.7)	212.9(2.2)	228.3(1.6)	242.9(2.3)	251.9(3.0)
SOME EDUC AFTER H.S.	226.5(1.6)	27.2(1.0)	179.4(2.4)	192.5(3.6)	209.0(2.0)	227.9(1.2)	245.2(1.5)	259.7(2.5)	269.0(4.0)
GRADUATED COLLEGE	222.9(0.9)	28.6(0.4)	172.8(2.0)	184.3(1.6)	204.0(1.2)	224.5(1.2)	243.3(1.4)	258.7(0.9)	267.2(2.0)
UNKNOWN	209.7(0.9)	26.6(0.5)	165.0(1.7)	175.1(1.4)	191.9(1.9)	210.7(1.3)	228.3(0.9)	242.7(1.5)	251.6(2.6)

428

429

BEST COPY AVAILABLE

Table 16-69

Weighted Proficiency Means, Standard Deviations, and Percentiles (with Standard Errors) for Main Focused-BIB Mathematics Samples, by Subgroup, Grade 8

	MEAN	STD DEV	5TH	10TH	25TH	50TH	75TH	90TH	95TH
-- TOTAL --	265.0(1.0)	32.4(0.6)	211.1(1.7)	222.8(1.0)	242.4(1.0)	265.8(1.1)	287.7(1.5)	306.5(1.2)	317.1(1.7)
GENDER									
MALE	265.5(1.3)	33.4(0.6)	211.3(2.0)	222.5(0.9)	241.8(1.7)	265.3(1.3)	289.5(2.4)	309.4(1.9)	320.4(2.8)
FEMALE	264.4(1.1)	31.3(0.7)	210.8(1.9)	223.0(1.2)	243.1(1.2)	266.2(1.2)	286.2(1.3)	303.7(1.6)	314.4(2.2)
RACE/ETHNICITY									
WHITE	272.1(1.2)	29.5(0.5)	222.8(1.2)	233.8(1.1)	251.8(1.5)	272.7(1.1)	292.4(1.3)	310.6(1.4)	320.3(2.1)
BLACK	240.8(1.6)	29.6(1.0)	192.3(2.9)	203.8(2.1)	221.1(2.6)	239.3(1.6)	259.7(1.5)	281.0(1.7)	293.2(5.0)
HISPANIC	247.9(1.6)	29.6(0.8)	199.5(2.3)	210.1(2.5)	227.4(2.1)	247.9(2.1)	268.1(1.4)	287.5(2.6)	297.8(3.4)
ASIAN/PACIFIC AMERICAN	284.8(4.1)	33.4(2.6)	227.3(8.9)	241.1(3.8)	263.9(4.6)	285.6(7.0)	307.3(2.8)	328.0(5.6)	339.5(4.7)
AMER IND/ALASKAN NATV	247.9(3.4)	28.9(3.0)	202.9(16.3)	210.6(6.8)	225.5(4.3)	247.5(7.4)	268.4(2.1)	286.7(5.5)	294.8(9.6)
PARENTS' EDUCATION LEVEL									
LESS THAN H.S.	245.8(1.4)	27.3(1.2)	202.5(1.7)	213.3(3.7)	228.0(1.6)	244.1(2.2)	263.2(2.2)	281.5(3.4)	293.3(12.6)
GRADUATED H.S.	255.5(0.9)	29.1(0.6)	207.2(1.2)	218.2(2.8)	236.0(1.2)	255.5(1.3)	275.4(0.8)	292.6(2.3)	302.9(2.3)
SOME EDUC AFTER H.S.	270.2(1.0)	28.9(0.7)	219.9(2.4)	231.8(4.3)	251.7(1.2)	271.6(1.6)	289.9(1.5)	305.9(1.0)	315.9(2.6)
GRADUATED COLLEGE	277.2(1.5)	30.9(0.5)	223.3(1.5)	236.1(1.9)	256.9(2.1)	278.8(1.4)	298.9(2.0)	315.9(1.8)	325.4(2.5)
UNKNOWN	242.6(1.9)	29.6(1.1)	193.5(4.4)	205.4(2.2)	222.3(3.3)	242.1(2.3)	263.1(4.0)	282.6(2.6)	293.5(2.8)

430 431

Table 16-70

Weighted Proficiency Means, Standard Deviations, and Percentiles (with Standard Errors)
for Main Focused-BIB Mathematics Samples, by Subgroup, Grade 12

	MEAN	STD DEV	5TH	10TH	25TH	50TH	75TH	90TH	95TH
-- TOTAL --	295.3(1.1)	33.3(0.5)	241.0(1.5)	250.8(1.0)	270.8(1.2)	295.9(1.3)	319.6(1.2)	339.3(1.3)	349.2(1.4)
GENDER									
MALE	297.7(1.3)	33.9(0.6)	243.3(2.0)	252.9(0.9)	272.0(1.6)	297.7(2.0)	323.0(1.7)	342.8(1.9)	353.6(2.6)
FEMALE	293.1(1.1)	32.6(0.5)	238.6(0.8)	248.9(1.5)	269.7(1.7)	294.5(1.5)	317.0(1.0)	335.1(1.4)	345.3(1.6)
RACE/ETHNICITY									
WHITE	301.1(1.2)	31.3(0.5)	249.0(1.5)	259.2(1.7)	278.5(1.6)	301.6(1.4)	323.7(1.2)	341.9(1.2)	351.6(2.2)
BLACK	270.2(1.3)	28.6(0.7)	226.0(1.8)	234.5(1.8)	248.9(1.3)	268.7(2.2)	289.6(3.2)	309.4(5.1)	320.2(2.9)
HISPANIC	277.6(2.4)	32.0(1.2)	228.5(5.9)	238.7(3.9)	254.0(2.5)	274.9(1.8)	300.3(4.2)	320.4(3.0)	332.2(7.0)
ASIAN/PACIFIC AMERICAN	315.0(4.0)	31.9(1.5)	257.3(4.2)	268.9(10.0)	293.8(11.5)	318.3(3.9)	337.9(10.4)	355.4(4.1)	365.0(6.5)
AMER IND/ALASKAN NATV	290.4(5.4)	31.3(3.9)	244.7(10.8)	254.0(7.7)	267.5(4.6)	290.3(6.5)	316.0(3.3)	331.9(5.4)	339.7(9.5)
PARENTS' EDUCATION LEVEL									
LESS THAN H.S.	272.4(1.4)	28.3(1.3)	228.4(3.7)	237.5(3.6)	251.8(1.3)	271.0(4.5)	290.0(3.1)	310.9(5.2)	322.7(4.1)
GRADUATED H.S.	282.2(1.2)	30.2(0.6)	233.8(3.6)	243.3(1.5)	260.5(1.1)	281.2(1.7)	303.4(1.3)	322.4(2.5)	333.7(2.9)
SOME EDUC AFTER H.S.	296.9(0.9)	29.8(0.5)	246.5(2.5)	255.9(1.7)	277.1(1.0)	298.4(1.7)	318.2(0.8)	335.6(1.5)	344.9(2.3)
GRADUATED COLLEGE	308.1(1.3)	31.9(0.6)	252.6(2.1)	263.5(2.4)	286.5(2.2)	309.7(1.4)	331.4(1.9)	348.2(1.8)	358.6(1.8)
UNKNOWN	268.0(3.5)	31.1(3.1)	220.0(7.9)	231.9(3.4)	247.2(3.7)	265.6(3.2)	286.2(10.2)	303.7(11.4)	326.8(21.7)

432

433

Table 16-71

Weighted Proficiency Means, Standard Deviations, and Percentiles (with Standard Errors) for Bridge to 1986 Mathematics Samples, by Subgroup, Age 9

	MEAN	STD DEV	5TH	10TH	25TH	50TH	75TH	90TH	95TH
-- TOTAL --	229.6(0.8)	32.9(0.5)	173.3(2.6)	185.8(2.2)	207.8(1.3)	231.1(0.9)	252.5(0.7)	271.0(1.0)	282.1(1.3)
GENDER									
MALE	229.1(0.9)	33.5(0.6)	171.8(2.5)	184.6(2.1)	206.7(1.2)	230.4(1.0)	252.4(0.8)	271.6(1.8)	282.8(1.7)
FEMALE	230.2(1.1)	32.4(0.6)	174.5(2.8)	187.0(2.7)	208.9(1.3)	231.8(1.0)	252.7(1.0)	270.4(1.3)	281.4(1.1)
OBSERVED RACE/ETHNICITY									
WHITE	235.2(0.8)	31.2(0.5)	181.8(2.4)	194.0(1.6)	214.6(0.9)	236.3(1.0)	256.4(0.6)	274.5(0.8)	284.8(2.1)
BLACK	208.4(2.2)	31.5(0.8)	156.0(1.7)	167.1(3.7)	186.0(4.1)	208.4(3.1)	231.4(2.1)	248.9(2.9)	258.9(4.3)
HISPANIC	213.8(2.1)	30.3(1.2)	161.8(3.4)	173.4(1.4)	193.1(3.6)	216.2(4.1)	235.1(3.3)	251.7(3.4)	262.2(3.5)
ASIAN/PACIFIC AMERICAN	236.2(3.3)	31.9(3.0)	183.8(3.6)	196.1(8.0)	216.1(8.5)	236.4(2.7)	258.7(3.1)	278.3(3.8)	287.3(5.6)
AMER IND/ALASKAN NATV	217.0(7.3)	28.7(4.7)	145.2(17.2)	187.9(24.3)	201.7(13.0)	217.3(9.6)	238.8(6.6)	247.7(11.4)	261.0(16.6)
PARENTS' EDUCATION LEVEL									
LESS THAN H.S.	210.4(2.3)	30.6(1.2)	158.7(4.2)	169.7(4.0)	188.4(1.8)	211.2(3.5)	233.1(3.2)	250.1(4.0)	259.4(5.2)
GRADUATED H.S.	226.2(1.2)	32.2(1.1)	168.7(8.2)	181.9(2.2)	205.5(1.0)	228.8(1.3)	248.5(1.2)	265.7(2.0)	275.7(4.6)
SOME EDUC AFTER H.S.	235.8(2.0)	32.6(2.0)	178.3(5.2)	191.3(4.2)	215.6(3.1)	238.5(2.8)	259.7(3.2)	276.3(4.2)	285.8(4.7)
GRADUATED COLLEGE	237.6(1.3)	32.5(0.9)	182.0(2.8)	195.0(1.8)	216.1(1.5)	238.9(1.8)	260.1(0.9)	278.4(1.8)	288.8(2.2)
UNKNOWN	223.0(1.0)	31.2(0.6)	170.2(2.0)	181.7(1.7)	202.3(1.3)	224.0(1.8)	244.9(1.4)	262.1(1.4)	273.2(1.9)

Table 16-72

Weighted Proficiency Means, Standard Deviations, and Percentiles (with Standard Errors)
for Bridge to 1986 Mathematics Samples, by Subgroup, Age 13

	MEAN	STD DEV	5TH	10TH	25TH	50TH	75TH	90TH	95TH
-- TOTAL --	270.4(0.9),	31.1(0.5)	217.6(2.2)	230.2(1.4)	249.8(0.9)	270.9(1.0)	291.7(1.0)	309.9(1.0)	320.1(1.6)
GENDER									
MALE	271.2(1.2)	32.4(0.7)	215.5(2.1)	228.6(2.0)	250.2(1.7)	272.0(1.0)	293.1(1.2)	312.4(1.4)	323.1(1.9)
FEMALE	269.6(0.9)	29.7(0.5)	220.4(2.3)	231.4(1.2)	249.5(1.1)	269.9(1.2)	290.3(1.3)	307.7(1.5)	317.3(0.8)
OBSERVED RACE/ETHNICITY									
WHITE	276.3(1.1)	29.0(0.5)	228.2(1.5)	239.3(1.0)	257.3(1.1)	276.6(1.0)	296.0(1.1)	313.2(1.3)	322.9(1.6)
BLACK	249.1(2.3)	28.7(1.2)	201.6(5.4)	211.8(2.2)	229.9(3.0)	249.4(2.0)	267.8(2.9)	285.3(2.8)	296.2(4.1)
HISPANIC	254.6(1.8)	29.9(1.2)	206.2(3.7)	216.4(3.1)	234.3(2.2)	255.1(1.9)	275.2(3.5)	292.2(2.9)	303.3(3.3)
ASIAN/PACIFIC AMERICAN	273.7(9.9)	33.1(2.1)	217.6(4.6)	229.8(8.0)	251.5(11.1)	273.7(14.1)	298.2(12.9)	316.2(12.7)	327.6(13.6)
AMER IND/ALASKAN NATV	265.1(7.9)	31.4(6.6)	188.3(17.7)	226.5(44.0)	250.1(15.0)	269.6(5.5)	285.2(2.4)	303.8(17.4)	310.6(15.7)
PARENTS' EDUCATION LEVEL									
LESS THAN H.S.	253.4(1.8)	28.3(1.3)	206.5(9.0)	216.7(1.4)	234.4(1.5)	254.4(2.2)	272.9(2.6)	288.5(2.4)	298.4(6.1)
GRADUATED H.S.	262.6(1.2)	27.5(0.7)	215.6(3.3)	226.6(2.5)	244.8(1.0)	262.9(2.0)	280.9(1.7)	297.8(1.8)	307.7(1.7)
SOME EDUC AFTER H.S.	277.1(1.0)	26.9(0.8)	232.6(1.9)	241.8(1.3)	260.1(1.3)	277.6(1.9)	294.8(1.7)	310.9(2.2)	319.8(2.2)
GRADUATED COLLEGE	280.4(1.0)	30.5(0.7)	227.5(1.7)	240.6(1.9)	260.6(1.8)	281.3(1.0)	301.7(1.4)	318.4(1.2)	327.8(2.0)
UNKNOWN	247.8(2.1)	30.0(1.3)	198.0(7.5)	208.9(2.4)	228.3(1.9)	247.1(2.4)	267.7(2.5)	286.5(5.4)	297.5(3.8)

436

437

Table 16-73

Weighted Proficiency Means, Standard Deviations, and Percentiles (with Standard Errors) for Bridge to 1986 Mathematics Samples, by Subgroup, Age 17

	MEAN	STD DEV	5TH	10TH	25TH	50TH	75TH	90TH	95TH
-- TOTAL --	304.6(0.9)	31.1(0.6)	253.4(1.0)	264.0(1.1)	282.5(1.0)	304.9(1.1)	326.5(1.2)	344.5(1.3)	355.5(2.2)
GENDER									
MALE	306.3(1.1)	32.3(0.7)	252.8(3.0)	263.9(1.2)	283.7(1.3)	306.4(1.6)	329.3(1.1)	347.8(1.4)	358.5(1.3)
FEMALE	302.9(1.1)	29.9(0.9)	253.9(1.9)	264.0(1.5)	281.5(1.3)	303.7(1.7)	324.1(1.2)	341.4(1.6)	351.8(2.2)
OBSERVED RACE/ETHNICITY									
WHITE	309.5(1.0)	29.5(0.5)	260.2(1.3)	270.5(1.5)	288.8(1.5)	310.1(1.3)	330.1(1.2)	347.2(1.0)	357.1(1.3)
BLACK	288.5(2.8)	27.9(1.7)	245.4(4.4)	253.5(3.5)	268.7(1.8)	287.1(2.5)	307.1(5.3)	325.7(5.8)	337.7(4.2)
HISPANIC	283.5(2.9)	31.8(1.8)	229.1(5.4)	242.2(8.1)	263.8(6.8)	281.8(2.4)	304.0(4.4)	325.1(3.6)	336.3(8.6)
ASIAN/PACIFIC AMERICAN	310.6(7.0)	34.9(3.3)	253.3(10.4)	266.0(5.7)	283.4(10.3)	310.3(6.6)	336.1(9.7)	358.5(7.5)	368.4(11.4)
AMER IND/ALASKAN NATV	307.3(7.1)	28.5(4.5)	258.8(16.9)	266.8(11.3)	293.4(17.2)	307.0(4.3)	321.4(19.9)	348.4(27.2)	355.0(14.9)
PARENTS' EDUCATION LEVEL									
LESS THAN H.S.	285.4(2.2)	27.9(1.2)	243.0(5.5)	249.9(3.2)	264.6(1.9)	283.4(2.2)	304.9(2.0)	323.3(3.1)	332.7(2.2)
GRADUATED H.S.	293.7(0.9)	28.4(0.8)	247.5(2.4)	257.3(1.5)	273.8(2.1)	293.5(2.3)	313.4(1.0)	330.0(1.9)	340.4(2.4)
SOME EDUC AFTER H.S.	307.7(1.0)	27.9(0.9)	262.5(2.0)	271.3(1.9)	287.7(1.1)	307.4(1.2)	326.5(1.4)	342.9(2.8)	354.6(2.3)
GRADUATED COLLEGE	316.2(1.3)	29.4(0.7)	267.0(2.3)	276.9(1.5)	296.3(1.3)	317.3(1.4)	337.3(2.1)	353.3(1.4)	362.3(1.3)
UNKNOWN	276.8(2.8)	31.3(2.2)	225.8(18.3)	234.3(7.7)	255.3(7.3)	275.7(6.0)	297.5(14.1)	317.5(7.8)	329.7(11.4)

Table 16-74

Weighted Proficiency Means, Standard Deviations, and Percentiles (with Standard Errors) for Main Focused-BIB Science Samples, by Subgroup, Grade 4

	MEAN	STD DEV	5TH	10TH	25TH	50TH	75TH	90TH	95TH
-- TOTAL --	232.8(0.9)	31.3(0.4)	179.4(1.3)	191.3(1.2)	211.6(1.1)	233.9(1.5)	254.9(1.2)	272.1(1.1)	282.0(1.4)
GENDER									
MALE	233.8(1.1)	31.8(0.6)	179.5(2.1)	191.4(1.5)	211.9(1.3)	235.0(1.5)	256.6(1.2)	274.0(1.4)	284.3(1.7)
FEMALE	231.7(1.0)	30.6(0.6)	179.2(1.4)	191.0(1.2)	211.3(1.2)	232.9(0.9)	253.0(1.3)	269.9(1.0)	279.6(2.5)
RACE/ETHNICITY									
WHITE	242.1(1.0)	27.3(0.6)	195.5(1.7)	206.3(1.8)	224.2(1.6)	242.8(1.0)	261.1(1.1)	276.5(0.9)	286.2(1.0)
BLACK	205.4(1.5)	26.6(0.7)	162.3(1.8)	172.2(2.2)	187.9(2.3)	204.8(1.6)	222.6(1.4)	239.2(1.8)	251.1(2.3)
HISPANIC	212.0(1.5)	29.2(1.0)	163.4(2.4)	173.9(5.1)	192.3(1.7)	211.7(1.9)	231.4(0.9)	250.0(1.7)	260.9(3.3)
ASIAN/PACIFIC AMERICAN	233.2(3.0)	30.1(2.2)	186.1(4.2)	195.2(7.9)	212.3(3.1)	230.7(4.9)	253.1(6.8)	274.9(6.2)	286.9(6.4)
AMER IND/ALASKAN NATV	226.1(2.7)	27.7(1.3)	180.5(6.9)	191.2(2.6)	207.0(6.2)	226.9(9.9)	245.8(4.5)	262.8(6.9)	270.1(4.9)
PARENTS' EDUCATION LEVEL									
LESS THAN H.S.	221.4(2.2)	27.9(1.5)	175.1(8.8)	185.1(4.6)	202.9(2.8)	222.0(2.7)	240.5(6.0)	258.1(6.1)	267.8(5.4)
GRADUATED H.S.	225.7(1.4)	28.9(0.8)	176.8(1.8)	188.3(2.0)	206.1(1.1)	226.4(1.6)	246.1(1.4)	262.9(2.0)	271.5(2.2)
SOME EDUC AFTER H.S.	241.9(1.8)	29.2(0.9)	187.9(3.6)	201.8(1.4)	223.7(3.9)	244.9(3.4)	262.4(1.8)	276.9(1.5)	284.8(3.6)
GRADUATED COLLEGE	242.7(1.2)	31.5(0.6)	186.8(2.0)	199.8(2.6)	221.9(2.1)	245.2(1.6)	265.5(1.1)	281.2(2.3)	290.6(1.1)
UNKNOWN	225.8(0.8)	29.4(0.6)	175.1(1.8)	187.4(2.9)	206.1(1.3)	227.2(0.9)	246.1(0.9)	262.5(2.4)	272.4(1.8)

440

441

Table 16-75

Weighted Proficiency Means, Standard Deviations, and Percentiles (with Standard Errors) for Main Focused-BIB Science Samples, by Subgroup, Grade 8

	MEAN	STD DEV	5TH	10TH	25TH	50TH	75TH	90TH	95TH
-- TOTAL --	263.1(1.2)	39.5(0.6)	195.3(2.2)	210.3(1.8)	236.6(1.4)	265.3(1.3)	291.0(1.5)	312.5(1.8)	324.4(2.2)
GENDER									
MALE	265.1(1.6)	41.3(0.8)	195.1(2.2)	210.2(2.9)	237.0(1.8)	266.9(2.0)	294.9(2.4)	317.7(1.5)	329.8(1.0)
FEMALE	261.0(1.2)	37.6(0.6)	195.5(1.7)	210.5(1.7)	236.0(1.3)	263.8(1.5)	287.8(1.9)	306.9(1.4)	318.3(1.9)
RACE/ETHNICITY									
WHITE	272.9(1.4)	35.7(0.6)	211.5(2.4)	225.4(1.9)	249.0(1.7)	274.7(1.8)	297.9(1.6)	317.6(1.3)	329.1(2.1)
BLACK	231.0(2.2)	36.5(1.2)	171.9(6.8)	185.3(2.6)	205.7(2.4)	230.2(2.8)	257.3(3.1)	277.8(2.7)	288.9(2.6)
HISPANIC	241.2(2.1)	36.2(1.0)	180.6(1.9)	193.6(3.2)	216.4(3.8)	242.4(3.7)	266.3(1.9)	286.7(2.1)	299.7(2.7)
ASIAN/PACIFIC AMERICAN	270.5(4.0)	36.5(2.2)	207.4(10.4)	221.6(4.6)	246.1(7.4)	270.8(8.8)	297.2(5.0)	317.0(5.9)	327.1(4.6)
AMER IND/ALASKAN NATV	251.9(8.5)	35.3(3.0)	190.4(8.7)	207.4(11.4)	230.1(12.2)	253.1(13.3)	279.1(7.3)	295.5(6.7)	307.5(5.5)
PARENTS' EDUCATION LEVEL									
LESS THAN H.S.	241.1(2.3)	33.7(1.4)	186.6(4.8)	197.8(5.2)	216.7(6.7)	242.1(5.7)	265.1(2.6)	283.3(2.3)	294.3(5.1)
GRADUATED H.S.	254.1(1.3)	36.7(1.0)	191.1(3.7)	205.0(2.3)	229.5(1.5)	256.2(1.2)	280.5(1.6)	299.9(1.6)	310.5(1.7)
SOME EDUC AFTER H.S.	268.4(1.4)	36.2(1.1)	205.0(7.7)	221.6(2.3)	246.5(2.0)	270.5(1.4)	293.1(2.0)	313.0(4.0)	324.2(5.4)
GRADUATED COLLEGE	276.4(1.7)	37.9(0.7)	209.3(2.4)	224.5(2.8)	251.6(2.7)	279.8(1.8)	303.5(1.7)	322.5(1.6)	333.7(2.5)
UNKNOWN	236.5(2.3)	38.3(1.4)	172.5(5.5)	186.3(6.0)	210.9(5.9)	236.9(4.1)	262.7(8.1)	285.8(7.5)	298.4(2.1)

442 443

Table 16-76

Weighted Proficiency Means, Standard Deviations, and Percentiles (with Standard Errors) for Main Focused-BIB Science Samples, by Subgroup, Grade 12

	MEAN	STD DEV	5TH	10TH	25TH	50TH	75TH	90TH	95TH
-- TOTAL --	293.5(1.2)	42.5(0.6)	221.7(1.8)	237.5(1.9)	265.0(1.6)	294.1(1.2)	323.2(1.1)	348.5(1.7)	362.4(1.7)
GENDER									
MALE	298.9(1.5)	43.9(0.7)	224.2(3.7)	240.6(2.1)	268.8(2.6)	299.3(1.6)	330.4(1.8)	356.4(1.7)	369.5(1.6)
FEMALE	288.6(1.2)	40.6(0.8)	219.7(2.4)	235.1(1.9)	261.7(1.8)	290.0(1.7)	316.8(1.8)	339.7(1.5)	353.1(2.8)
RACE/ETHNICITY									
WHITE	302.5(1.3)	38.8(0.6)	237.7(2.4)	251.5(1.9)	276.6(1.5)	302.6(1.5)	329.4(1.2)	353.1(2.6)	365.8(2.2)
BLACK	256.3(2.4)	38.5(1.7)	194.8(3.6)	208.1(3.0)	229.4(2.1)	256.5(1.9)	281.8(2.6)	304.3(2.8)	319.6(4.1)
HISPANIC	272.5(2.8)	38.7(1.6)	211.0(5.7)	223.1(6.3)	245.5(3.2)	272.3(4.2)	297.4(3.8)	322.9(3.8)	337.7(7.5)
ASIAN/PACIFIC AMERICAN	308.2(7.1)	42.5(2.1)	230.2(21.6)	249.6(8.4)	280.3(14.7)	311.0(10.0)	338.8(9.9)	360.9(10.3)	372.2(11.9)
AMER IND/ALASKAN NATV	285.7(4.6)	31.6(4.2)	236.8(7.2)	246.6(12.3)	262.8(10.5)	284.7(24.9)	308.0(11.1)	329.7(9.7)	344.5(14.9)
PARENTS' EDUCATION LEVEL									
LESS THAN H.S.	269.0(2.5)	37.8(1.6)	209.1(5.7)	222.5(3.0)	242.8(3.4)	266.9(3.9)	294.8(9.4)	319.1(4.3)	334.2(4.7)
GRADUATED H.S.	278.9(1.3)	36.6(0.8)	214.5(2.8)	228.9(2.4)	252.7(2.9)	280.0(1.9)	305.7(1.3)	328.2(1.9)	341.5(3.5)
SOME EDUC AFTER H.S.	295.2(1.3)	38.3(0.9)	229.4(3.4)	245.6(3.5)	270.4(1.8)	295.5(1.3)	321.6(1.5)	344.6(2.3)	357.7(5.1)
GRADUATED COLLEGE	308.2(1.4)	41.3(0.8)	236.3(2.5)	252.9(1.8)	281.2(1.0)	309.8(1.5)	337.8(2.2)	360.5(3.1)	372.3(1.5)
UNKNOWN	249.0(5.3)	44.0(4.7)	177.9(8.2)	194.1(12.4)	216.8(7.8)	248.7(7.0)	279.2(12.0)	301.7(4.5)	314.7(7.8)

444 445

Table 16-77

Weighted Proficiency Means, Standard Deviations, and Percentiles (with Standard Errors) for Bridge to 1986 Science Samples, by Subgroup, Age 9

	MEAN	STD DEV	5TH	10TH	25TH	50TH	75TH	90TH	95TH
-- TOTAL --	228.7(0.8)	40.2(0.4)	159.8(1.3)	176.1(1.1)	202.0(1.4)	230.3(0.9)	256.6(0.8)	278.8(1.3)	292.1(1.4)
GENDER									
MALE	230.3(1.1)	41.9(0.6)	159.6(2.2)	176.3(2.3)	202.1(2.5)	231.6(1.9)	259.4(1.0)	283.3(1.8)	296.3(2.4)
FEMALE	227.1(1.0)	38.4(0.5)	159.9(2.4)	175.8(2.2)	201.9(1.2)	229.2(1.1)	254.0(1.1)	274.6(1.9)	287.0(1.9)
OBSERVED RACE/ETHNICITY									
WHITE	237.5(0.8)	36.3(0.4)	176.9(1.4)	189.9(1.3)	212.6(0.8)	238.3(1.0)	262.3(1.0)	283.5(1.4)	295.7(1.3)
BLACK	196.4(2.0)	38.6(1.0)	131.3(4.2)	145.3(3.8)	169.8(2.6)	196.3(2.5)	224.1(1.7)	246.8(2.4)	260.0(5.4)
HISPANIC	206.2(2.2)	37.0(1.7)	146.2(5.5)	158.6(4.3)	180.6(3.7)	206.2(3.7)	232.7(4.1)	252.9(4.4)	266.8(6.9)
ASIAN/PACIFIC AMERICAN	226.7(4.0)	40.3(5.7)	156.0(8.8)	173.8(11.5)	201.1(14.2)	227.6(3.2)	253.7(10.3)	275.8(10.8)	291.7(5.5)
AMER IND/ALASKAN NATV	222.4(16.3)	42.5(8.9)	150.3(17.4)	167.3(17.0)	186.3(14.8)	220.2(10.6)	250.7(16.1)	291.9(70.3)	295.2(42.4)
PARENTS' EDUCATION LEVEL									
LESS THAN H.S.	209.8(2.7)	39.4(1.9)	143.3(8.0)	158.1(7.8)	182.1(3.3)	210.2(3.2)	236.7(3.2)	261.6(7.7)	275.6(4.7)
GRADUATED H.S.	225.8(1.7)	38.5(1.1)	159.7(7.2)	174.6(5.0)	200.3(3.5)	227.4(3.4)	252.7(2.3)	274.0(1.6)	285.6(3.4)
SOME EDUC AFTER H.S.	237.6(2.1)	41.5(1.6)	165.2(20.6)	180.7(7.9)	210.8(2.6)	241.8(2.6)	266.3(2.5)	287.3(3.0)	300.5(5.8)
GRADUATED COLLEGE	236.2(1.3)	39.7(0.9)	168.1(2.7)	183.8(2.0)	210.2(1.6)	238.3(1.2)	263.2(1.6)	285.8(1.5)	298.1(1.3)
UNKNOWN	221.5(1.2)	39.1(0.7)	154.8(1.7)	171.1(1.0)	195.5(2.6)	222.8(2.0)	248.7(2.1)	270.9(2.3)	283.7(1.8)

446

447

Table 16-78

Weighted Proficiency Means, Standard Deviations, and Percentiles (with Standard Errors) for Bridge to 1986 Science Samples, by Subgroup, Age 13

	MEAN	STD DEV	5TH	10TH	25TH	50TH	75TH	90TH	95TH
-- TOTAL --	255.2(0.9)	37.6(0.7)	191.4(2.0)	205.9(1.7)	230.0(1.5)	256.4(1.2)	281.1(0.9)	302.4(1.1)	315.1(1.9)
GENDER									
MALE	258.5(1.1)	38.8(0.8)	191.9(2.5)	207.3(3.4)	232.9(1.4)	260.3(1.4)	285.8(2.2)	307.4(1.5)	320.2(1.2)
FEMALE	251.8(1.1)	36.1(0.8)	190.6(2.1)	204.8(1.5)	227.8(1.6)	253.1(1.2)	276.8(1.6)	296.8(1.1)	308.6(1.4)
OBSERVED RACE/ETHNICITY									
WHITE	264.1(0.9)	33.8(0.5)	208.6(1.6)	220.4(1.2)	241.3(0.9)	264.5(1.1)	287.0(1.7)	307.1(1.4)	319.4(1.3)
BLACK	225.7(3.1)	34.3(1.7)	169.7(5.5)	181.8(6.1)	202.3(3.7)	225.7(3.0)	249.1(2.6)	269.0(4.2)	283.2(3.7)
HISPANIC	231.6(2.6)	36.6(1.0)	173.7(4.7)	185.3(4.5)	205.9(4.1)	230.9(3.3)	256.4(5.1)	280.0(5.9)	294.2(2.8)
ASIAN/PACIFIC AMERICAN	246.6(14.1)	40.4(2.0)	181.3(19.8)	193.2(21.9)	219.9(15.9)	246.4(13.2)	272.8(11.8)	297.7(19.8)	312.5(32.0)
AMER IND/ALASKAN NATV	251.5(11.2)	35.1(7.0)	181.6(22.0)	206.2(82.9)	227.2(13.6)	254.6(13.5)	278.7(13.1)	296.3(14.9)	301.8(14.8)
PARENTS' EDUCATION LEVEL									
LESS THAN H.S.	232.9(2.1)	34.4(1.3)	176.9(8.2)	189.6(3.4)	209.5(3.1)	232.0(1.7)	256.6(2.3)	276.5(6.0)	289.9(3.6)
GRADUATED H.S.	247.3(1.3)	34.5(0.9)	190.2(2.6)	203.1(2.7)	224.5(1.0)	247.9(1.7)	270.3(1.9)	291.8(2.2)	304.3(2.6)
SOME EDUC AFTER H.S.	262.8(1.2)	33.2(0.9)	206.8(2.0)	219.7(3.0)	240.7(1.9)	263.4(1.2)	285.5(1.4)	304.4(1.6)	315.9(2.4)
GRADUATED COLLEGE	267.5(1.1)	35.5(0.9)	206.2(3.5)	220.8(1.7)	244.3(1.3)	268.9(1.4)	291.6(1.4)	311.7(1.3)	323.8(1.5)
UNKNOWN	224.3(2.1)	35.6(1.6)	164.3(6.4)	177.8(4.2)	200.1(2.1)	224.1(1.8)	248.2(1.6)	270.0(4.7)	284.7(6.5)

448

449

Table 16-79

Weighted Proficiency Means, Standard Deviations, and Percentiles (with Standard Errors) for Bridge to 1986 Science Samples, by Subgroup, Age 17

	MEAN	STD DEV	5TH	10TH	25TH	50TH	75TH	90TH	95TH
-- TOTAL --	290.4(1.1)	46.2(0.6)	209.9(2.3)	228.8(2.0)	260.3(1.9)	292.2(1.3)	322.7(1.4)	348.3(1.2)	362.9(1.5)
GENDER									
MALE	295.6(1.3)	48.7(0.9)	210.4(3.9)	229.5(2.9)	263.4(1.3)	297.9(1.9)	329.9(1.8)	356.7(2.3)	372.5(1.8)
FEMALE	285.4(1.6)	43.2(1.0)	?09.2(3.7)	228.2(4.5)	257.7(2.4)	287.7(2.0)	316.2(2.3)	339.6(2.3)	351.5(1.6)
OBSERVED RACE/ETHNICITY									
WHITE	300.9(1.1)	41.1(0.6)	232.8(2.3)	249.0(2.0)	273.4(1.5)	301.2(1.2)	329.0(1.6)	352.3(1.3)	367.3(2.0)
BLACK	253.0(4.5)	44.7(2.4)	182.0(10.1)	196.6(3.1)	220.5(4.3)	251.6(3.0)	282.9(6.0)	313.6(11.3)	329.3(10.2)
HISPANIC	261.5(4.4)	44.1(2.6)	188.7(6.2)	203.9(11.1)	230.6(3.6)	260.5(5.7)	292.6(10.6)	317.4(5.1)	329.5(9.1)
ASIAN/PACIFIC AMERICAN	283.0(7.9)	49.3(3.9)	203.1(10.3)	223.4(22.8)	252.9(16.4)	285.4(14.2)	325.8(10.5)	356.6(21.4)	370.0(6.4)
AMER IND/ALASKAN NATV	280.8(12.1)	41.3(6.3)	208.6(15.1)	223.0(15.9)	252.2(27.1)	281.1(8.1)	311.4(18.9)	330.2(19.3)	347.7(25.9)
PARENTS' EDUCATION LEVEL									
LESS THAN H.S.	261.4(2.8)	41.9(2.2)	186.3(4.7)	206.3(10.2)	234.0(13.4)	262.5(4.3)	291.6(5.4)	315.3(9.2)	328.3(14.0)
GRADUATED H.S.	276.3(1.4)	43.9(1.3)	200.4(3.8)	215.9(2.8)	247.5(3.6)	278.9(2.4)	306.9(1.1)	331.7(2.1)	345.2(3.2)
SOME EDUC AFTER H.S.	296.5(1.6)	40.5(1.4)	230.0(5.7)	246.3(2.8)	269.5(2.0)	297.1(1.9)	322.9(2.3)	347.3(5.3)	362.2(3.4)
GRADUATED COLLEGE	305.5(1.7)	44.3(0.7)	227.5(4.1)	246.9(3.1)	276.3(2.3)	308.3(2.6)	337.1(2.1)	359.6(2.9)	373.9(2.3)
UNKNOWN	248.2(5.5)	45.7(4.4)	165.1(40.9)	191.3(9.1)	217.3(7.1)	247.8(27.9)	278.3(8.3)	311.9(13.3)	327.8(6.7)

450

451

Table 16-80

Weighted Response Percentages and Proficiency Means (with Standard Errors)
for Main Focused-BIB Reading Samples, by Gender, Grade 4

GENDER OF SUBJECT	N	WEIGHTED PCT <CV>	MALE	FEMALE	MISSING
-- TOTAL --	9041	100.0(0.0) < 1%>	50.0(0.6) 228.1(1.1)	50.0(0.6) 238.0(1.0)	0.0 0.0
GENDER					
MALE	4508	50.0(0.6) < 1%>	100.0(0.0) 228.1(1.1)	0.0(0.0) 0.0(0.0)	0.0 0.0
FEMALE	4533	50.0(0.6) < 1%>	0.0(0.0) 0.0(0.0)	100.0(0.0) 238.0(1.0)	0.0 0.0
RACE/ETHNICITY					
WHITE	5307	70.1(0.4) < 1%>	49.7(0.7) 236.8(1.2)	50.3(0.7) 245.5(1.1)	0.0 0.0
BLACK	1544	15.3(0.3) < 2%>	49.6(1.3) 203.0(2.2)	50.4(1.3) 215.8(2.3)	0.0 0.0
HISPANIC	1631	10.6(0.2) < 2%>	51.1(1.2) 206.7(1.9)	48.9(1.2) 219.5(2.1)	0.0 0.0
ASIAN/PACIFIC AMERICAN	319	2.2(0.3) <11%>	50.5(2.9) 231.0(4.1)	49.5(2.9) 242.3(3.5)	0.0 0.0
AMER IND/ALASKAN NATV	224	1.6(0.3) <16%>	57.0(3.5) 225.6(3.1)	43.0(3.5) 233.8(4.5)	0.0 0.0
PARENTS' EDUCATION LEVEL					
LESS THAN H.S.	424	4.9(0.3) < 7%>	47.9(2.5) 207.5(3.3)	52.1(2.5) 220.8(2.8)	0.0 0.0
GRADUATED H.S.	1314	15.9(0.6) < 4%>	49.7(1.5) 219.9(2.2)	50.3(1.5) 234.8(1.8)	0.0 0.0
SOME EDUC AFTER H.S.	750	8.5(0.3) < 4%>	48.5(2.5) 235.7(2.9)	51.5(2.5) 247.1(2.3)	0.0 0.0
GRADUATED COLLEGE	3416	36.3(0.9) < 3%>	53.0(1.0) 238.7(1.8)	47.0(1.0) 245.9(1.5)	0.0 0.0
UNKNOWN	3085	33.9(0.8) < 3%>	47.5(1.1) 221.1(1.4)	52.5(1.1) 232.2(1.2)	0.0 0.0

452 453

Table 16-81

Weighted Response Percentages and Proficiency Means (with Standard Errors)
for Main Focused-BIB Reading Samples, by Derived Race/Ethnicity, Grade 4

DERIVED RACE/ETHNICITY	N	WEIGHTED PCT <CV>	WHITE	BLACK	HISPANIC	ASIAN	AMER IND	UNCLASSI	MISSING
-- TOTAL --	9041	100.0(0.0) < 1%>	70.1(0.4) 241.2(0.9)	15.3(0.3) 209.4(1.9)	10.6(0.2) 212.9(1.6)	2.2(0.3) 236.6(3.0)	1.6(0.3) 229.1(2.5)	0.1(0.0) 247.1(11.6)	0.0 0.0
GENDER									
MALE	4508	50.0(0.6) < 1%>	69.7(0.6) 236.8(1.2)	15.2(0.5) 203.0(2.2)	10.9(0.3) 206.7(1.9)	2.3(0.3) 231.0(4.1)	1.9(0.3) 225.6(3.1)	0.1(0.1) 256.5(13.0)	0.0 0.0
FEMALE	4533	50.0(0.6) < 1%>	70.4(0.6) 245.5(1.1)	15.5(0.4) 215.8(2.3)	10.4(0.3) 219.5(2.1)	2.2(0.3) 242.3(3.5)	1.4(0.2) 233.8(4.5)	0.1(0.0) 226.9(18.2)	0.0 0.0
RACE/ETHNICITY									
WHITE	5307	70.1(0.4) < 1%>	100.0(0.0) 241.2(0.9)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
BLACK	1544	15.3(0.3) < 2%>	0.0(0.0) 0.0(0.0)	100.0(0.0) 209.4(1.9)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
HISPANIC	1631	10.6(0.2) < 2%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 212.9(1.6)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
ASIAN/PACIFIC AMERICAN	319	2.2(0.3) <11%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 236.6(3.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
AMER IND/ALASKAN NATV	224	1.6(0.3) <16%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 229.1(2.5)	0.0(0.0) 0.0(0.0)	0.0 0.0
PARENTS' EDUCATION LEVEL									
LESS THAN H.S.	424	4.9(0.3) < 7%>	63.5(2.9) 221.0(3.0)	16.2(2.5) 198.0(5.6)	16.1(1.6) 201.5(4.6)	1.6(0.6) 227.1(15.8)	2.6(0.8) 226.8(11.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
GRADUATED H.S.	1314	15.9(0.6) < 4%>	70.3(1.4) 234.4(1.8)	16.7(1.2) 209.1(2.8)	10.4(0.8) 210.1(2.5)	0.9(0.3) 207.2(6.1)	1.6(0.5) 231.8(6.7)	0.0(0.0) 214.5(11.0)	0.0 0.0
SOME EDUC AFTER H.S.	750	8.5(0.3) < 4%>	73.9(1.8) 249.9(2.3)	13.5(1.6) 216.3(6.6)	9.4(1.0) 216.0(4.6)	1.9(0.5) 230.3(5.6)	1.1(0.6) 230.4(11.8)	0.2(0.2) 254.1(8.5)	0.0 0.0
GRADUATED COLLEGE	3416	36.3(0.9) < 3%>	70.8(0.9) 251.4(1.3)	16.4(0.7) 213.0(2.0)	8.1(0.4) 219.5(2.6)	2.9(0.4) 245.8(4.7)	1.7(0.3) 234.1(4.9)	0.2(0.1) 260.4(8.1)	0.0 0.0
UNKNOWN	3085	33.9(0.8) < 3%>	69.0(0.9) 234.0(1.2)	13.9(0.8) 206.1(3.0)	13.0(0.5) 211.2(2.1)	2.4(0.3) 232.4(3.9)	1.6(0.2) 222.7(4.3)	0.1(0.0) 202.4(28.2)	0.0 0.0

Table 16-82

Weighted Response Percentages and Proficiency Means (with Standard Errors)
for Main Focused-BIB Reading Samples, by Parents' Education Level, Grade 4

PARENTS' EDUCATION LEVEL	N	WEIGHTED PCT <CV>	NOT HS	GRAD HS	POST HS	GRAD COL	UNKNOWN	MISSING
-- TOTAL --	8989	100.0(0.0) < 1%>	4.9(0.3) 214.4(1.9)	16.0(0.6) 227.4(1.4)	8.5(0.3) 241.6(2.0)	36.5(0.9) 242.1(1.3)	34.1(0.8) 226.9(1.0)	0.5 204.0
GENDER MALE	4475	50.0(0.6) < 1%>	4.7(0.4) 207.5(3.3)	15.9(0.8) 219.9(2.2)	8.3(0.6) 235.7(2.9)	38.7(1.1) 238.7(1.8)	32.4(1.1) 221.1(1.4)	0.6 198.4
FEMALE	4514	50.0(0.6) < 1%>	5.1(0.5) 220.8(2.8)	16.1(0.7) 234.8(1.8)	8.8(0.5) 247.1(2.3)	34.3(1.0) 245.9(1.5)	35.8(0.8) 232.2(1.2)	0.4 211.7
RACE/ETHNICITY WHITE	5271	70.0(0.4) < 1%>	4.5(0.4) 221.0(3.0)	16.0(0.8) 234.4(1.8)	9.0(0.4) 249.9(2.3)	36.9(1.1) 251.4(1.3)	33.6(0.9) 234.0(1.2)	0.5 212.0
BLACK	1534	15.3(0.3) < 2%>	5.2(0.8) 198.0(5.6)	17.4(1.1) 209.1(2.8)	7.5(0.9) 216.3(6.6)	39.0(1.5) 213.0(2.0)	30.9(1.7) 206.1(3.0)	0.6 172.8
HISPANIC	1626	10.7(0.2) < 2%>	7.4(0.6) 201.5(4.6)	15.7(1.0) 210.1(2.5)	7.5(0.7) 216.0(4.6)	27.8(1.5) 219.5(2.6)	41.7(1.7) 211.2(2.1)	0.4 206.2
ASIAN/PACIFIC AMERICAN	318	2.2(0.3) <11%>	3.6(1.5) 227.1(15.8)	6.1(1.7) 207.2(6.1)	7.1(1.7) 230.3(5.6)	46.6(3.7) 245.8(4.7)	36.6(4.2) 232.4(3.9)	0.2 168.7
AMER IND/ALASKAN NATV	224	1.6(0.3) <16%>	7.7(2.1) 226.8(11.0)	16.0(3.0) 231.8(6.7)	5.9(2.2) 230.4(11.8)	36.9(3.9) 234.1(4.9)	33.5(3.2) 222.7(4.3)	0.0 0.0
PARENTS' EDUCATION LEVEL LESS THAN H.S.	424	4.9(0.3) < 7%>	100.0(0.0) 214.4(1.9)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
GRADUATED H.S.	1314	16.0(0.6) < 4%>	0.0(0.0) 0.0(0.0)	100.0(0.0) 227.4(1.4)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
SOME EDUC AFTER H.S.	750	8.5(0.3) < 4%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 241.6(2.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
GRADUATED COLLEGE	3416	36.5(0.9) < 3%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 242.1(1.3)	0.0(0.0) 0.0(0.0)	0.0 0.0
UNKNOWN	3085	34.1(0.8) < 3%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 226.9(1.0)	0.0 0.0

Table 16-83

Weighted Response Percentages and Proficiency Means (with Standard Errors)
for Main Focused-BIB Reading Samples, by Gender, Grade 8

GENDER OF SUBJECT	N	WEIGHTED PCT <CV>	MALE	FEMALE	MISSING
-- TOTAL --	6510	100.0(0.0) < 1%>	49.6(0.7) 252.7(1.2)	50.4(0.7) 268.3(1.0)	0.0 0.0
GENDER					
MALE	3220	49.6(0.7) < 2%>	100.0(0.0) 252.7(1.2)	0.0(0.0) 0.0(0.0)	0.0 0.0
FEMALE	3290	50.4(0.7) < 2%>	0.0(0.0) 0.0(0.0)	100.0(0.0) 268.3(1.0)	0.0 0.0
RACE/ETHNICITY					
WHITE	4176	70.2(0.4) < 1%>	50.1(0.8) 257.8(1.5)	49.9(0.8) 273.8(1.1)	0.0 0.0
BLACK	967	15.4(0.4) < 3%>	46.3(1.7) 238.1(2.1)	53.7(1.7) 253.6(2.0)	0.0 0.0
HISPANIC	968	10.0(0.3) < 3%>	53.0(1.8) 237.0(2.4)	47.0(1.8) 250.5(2.5)	0.0 0.0
ASIAN/PACIFIC AMERICAN	268	2.7(0.4) <14%>	45.8(4.0) 271.4(4.6)	54.2(4.0) 282.3(4.3)	0.0 0.0
AMER IND/ALASKAN NATV	119	1.7(0.4) <22%>!	46.1(5.8) 237.4(4.9)	53.9(5.8) 257.4(4.6)	0.0 0.0
PARENTS' EDUCATION LEVEL					
LESS THAN H.S.	570	9.0(0.6) < 7%>	44.4(2.2) 237.2(2.8)	55.6(2.2) 251.5(2.4)	0.0 0.0
GRADUATED H.S.	1579	25.7(0.9) < 4%>	50.9(1.2) 244.8(1.8)	49.1(1.2) 257.3(1.5)	0.0 0.0
SOME EDUC AFTER H.S.	1190	18.3(0.7) < 4%>	49.1(1.2) 255.9(2.0)	50.9(1.2) 273.9(1.7)	0.0 0.0
GRADUATED COLLEGE	2628	39.5(1.5) < 4%>	49.7(0.9) 262.9(1.5)	50.3(0.9) 280.5(1.3)	0.0 0.0
UNKNOWN	539	7.5(0.3) < 5%>	52.0(2.6) 237.2(2.6)	48.0(2.6) 248.0(3.0)	0.0 0.0

! INTERPRET WITH CAUTION: SAMPLING ERROR CANNOT BE ESTIMATED ACCURATELY SINCE COEFFICIENT OF VARIATION OF ESTIMATED NUMBER OF STUDENTS EXCEEDS 20%

Table 16-84

Weighted Response Percentages and Proficiency Means (with Standard Errors) for Main Focused-BIB Reading Samples, by Derived Race/Ethnicity, Grade 8

DERIVED RACE/ETHNICITY	N	WEIGHTED PCT <CV>	WHITE	BLACK	HISPANIC	ASIAN	AMER IND	UNCLASSI	MISSING
-- TOTAL --	6510	100.0(0.0) < 1%>	70.2(0.4) 265.8(1.1)	15.4(0.4) 246.4(1.6)	10.0(0.3) 243.3(1.8)	2.7(0.4) 277.3(3.4)	1.7(0.4) 248.2(3.3)	0.1(0.0) 252.6(13.6)	0.0 0.0
GENDER MALE	3220	49.6(0.7) < 2%>	70.9(0.8) 257.8(1.5)	14.4(0.6) 238.1(2.1)	10.7(0.4) 237.0(2.4)	2.4(0.4) 271.4(4.6)	1.6(0.3) 237.4(4.9)	0.1(0.0) 208.3(25.8)	0.0 0.0
FEMALE	3290	50.4(0.7) < 2%>	69.5(0.7) 273.8(1.1)	16.4(0.6) 253.6(2.0)	9.3(0.4) 250.5(2.5)	2.9(0.5) 282.3(4.3)	1.8(0.5) 257.4(4.6)	0.1(0.0) 270.0(16.0)	0.0 0.0
RACE/ETHNICITY WHITE	4176	70.2(0.4) < 1%>	100.0(0.0) 265.8(1.1)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
BLACK	967	15.4(0.4) < 3%>	0.0(0.0) 0.0(0.0)	100.0(0.0) 246.4(1.6)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
HISPANIC	968	10.0(0.3) < 3%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 243.3(1.8)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
ASIAN/PACIFIC AMERICAN	268	2.7(0.4) <14%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 277.3(3.4)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
AMER IND/ALASKAN NATV	119	1.7(0.4) <22%> !	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 248.2(3.3)	0.0(0.0) 0.0(0.0)	0.0 0.0
PARENTS' EDUCATION LEVEL LESS THAN H.S.	570	9.0(0.6) < 7%>	57.9(3.0) 247.9(2.8)	15.5(2.4) 239.5(4.7)	21.5(2.1) 244.1(2.9)	1.0(0.3) 240.9(13.4)	4.1(1.8) 233.8(7.1)	0.1(0.1) 267.1(****)	0.0 0.0
GRADUATED H.S.	1579	25.7(0.9) < 4%>	72.2(1.4) 254.9(1.5)	15.5(1.1) 237.5(2.4)	9.7(0.8) 243.2(2.6)	1.3(0.3) 255.7(9.6)	1.3(0.3) 245.0(5.6)	0.0(0.0) 0.0(0.0)	0.0 0.0
SOME EDUC AFTER H.S.	1190	18.3(0.7) < 4%>	72.4(1.5) 270.0(1.6)	15.1(1.4) 252.1(2.9)	8.6(0.8) 245.4(4.4)	1.6(0.4) 280.8(5.4)	2.2(0.5) 259.5(8.4)	0.1(0.1) 224.4(****)	0.0 0.0
GRADUATED COLLEGE	2628	39.5(1.5) < 4%>	75.5(1.0) 276.1(1.4)	13.9(0.8) 254.0(2.3)	5.7(0.5) 251.4(2.9)	3.7(0.5) 288.0(3.3)	1.0(0.2) 252.3(5.8)	0.1(0.1) 252.4(23.8)	0.0 0.0
UNKNOWN	539	7.5(0.3) < 5%>	45.0(2.3) 246.0(3.6)	23.1(1.9) 240.2(4.2)	23.0(1.6) 230.5(3.2)	6.3(1.6) 264.6(4.8)	2.3(0.9) 249.0(14.0)	0.2(0.2) 271.6(5.3)	0.0 0.0

! INTERPRET WITH CAUTION: SAMPLING ERROR CANNOT BE ESTIMATED ACCURATELY SINCE COEFFICIENT OF VARIATION OF ESTIMATED NUMBER OF STUDENTS EXCEEDS 20%

Table 16-85

Weighted Response Percentages and Proficiency Means (with Standard Errors)
for Main Focused-BIB Reading Samples, by Parents' Education Level, Grade 8

PARENTS' EDUCATION LEVEL	N	WEIGHTED PCT <CV>	NOT HS	GRAD HS	POST HS	GRAD COL	UNKNOWN	MISSING
-- TOTAL --	6506	100.0(0.0) < 1%>	9.0(0.6) 245.1(2.0)	25.7(0.9) 250.9(1.3)	18.3(0.7) 265.1(1.4)	39.5(1.5) 271.8(1.2)	7.5(0.3) 242.4(2.3)	0.1 216.6
GENDER								
MALE	3217	49.6(0.7) < 2%>	8.0(0.6) 237.2(2.8)	26.4(1.1) 244.8(1.8)	18.1(0.8) 255.9(2.0)	39.6(1.6) 262.9(1.5)	7.9(0.5) 237.2(2.6)	0.1 226.0
FEMALE	3289	50.4(0.7) < 2%>	9.9(0.8) 251.5(2.4)	25.0(1.0) 257.3(1.5)	18.5(0.7) 273.9(1.7)	39.4(1.6) 280.5(1.3)	7.2(0.5) 248.0(3.0)	0.0 158.2
RACE/ETHNICITY								
WHITE	4173	70.2(0.4) < 1%>	7.4(0.8) 247.9(2.8)	26.4(1.2) 254.9(1.5)	18.9(0.9) 270.0(1.6)	42.5(1.8) 276.1(1.4)	4.8(0.4) 246.0(3.6)	0.1 226.0
BLACK	966	15.4(0.4) < 3%>	9.0(1.5) 239.5(4.7)	25.9(1.9) 237.5(2.4)	18.0(1.6) 252.1(2.9)	35.8(2.3) 254.0(2.3)	11.3(1.0) 240.2(4.2)	0.1 158.2
HISPANIC	968	10.0(0.3) < 3%>	19.4(2.0) 244.1(2.9)	25.0(1.6) 243.2(2.6)	15.8(1.2) 245.4(4.4)	22.5(1.7) 251.4(2.9)	17.4(1.3) 230.5(3.2)	0.0 0.0
ASIAN/PACIFIC AMERICAN	268	2.7(0.4) <14%>	3.4(1.0) 240.9(13.4)	12.7(2.5) 255.7(9.6)	10.9(2.2) 280.8(5.4)	55.1(5.5) 288.0(3.3)	18.0(3.1) 264.6(4.8)	0.0 0.0
AMER IND/ALASKAN NATV	119	1.7(0.4) <22%>!	21.9(7.1) 233.8(7.1)	20.0(4.6) 245.0(5.6)	24.0(3.4) 259.5(8.4)	23.7(6.4) 252.3(5.8)	10.3(2.9) 249.0(14.0)	0.0 0.0
PARENTS' EDUCATION LEVEL								
LESS THAN H.S.	570	9.0(0.6) < 7%>	100.0(0.0) 245.1(2.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
GRADUATED H.S.	1579	25.7(0.9) < 4%>	0.0(0.0) 0.0(0.0)	100.0(0.0) 250.9(1.3)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
SOME EDUC AFTER H.S.	1190	18.3(0.7) < 4%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 265.1(1.4)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
GRADUATED COLLEGE	2628	39.5(1.5) < 4%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 271.8(1.2)	0.0(0.0) 0.0(0.0)	0.0 0.0
UNKNOWN	539	7.5(0.3) < 5%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 242.4(2.3)	0.0 0.0

! INTERPRET WITH CAUTION: SAMPLING ERROR CANNOT BE ESTIMATED ACCURATELY SINCE COEFFICIENT OF VARIATION OF ESTIMATED NUMBER OF STUDENTS EXCEEDS 20%

Table 16-36

Weighted Response Percentages and Proficiency Means (with Standard Errors)
for Main Focused-BIB Reading Samples, by Gender, Grade 12

GENDER OF SUBJECT	N	WEIGHTED PCT <CV>	MALE	FEMALE	MISSING
-- TOTAL --	6258	100.0(0.0) < 1%>	49.0(0.8) 283.3(1.3)	51.0(0.8) 293.4(0.9)	0.0 0.0
GENDER					
MALE	3022	49.0(0.8) < 2%>	100.0(0.0) 283.3(1.3)	0.0(0.0) 0.0(0.0)	0.0 0.0
FEMALE	3236	51.0(0.8) < 2%>	0.0(0.0) 0.0(0.0)	100.0(0.0) 293.4(0.9)	0.0 0.0
RACE/ETHNICITY					
WHITE	4453	73.5(0.7) < 2%>	49.8(0.9) 289.3(1.3)	50.2(0.9) 298.8(0.9)	0.0 0.0
BLACK	844	14.5(0.6) < 4%>	42.1(2.1) 262.4(2.4)	57.9(2.1) 274.1(2.1)	0.0 0.0
HISPANIC	666	8.0(0.2) < 3%>	51.2(2.4) 261.6(2.7)	48.8(2.4) 281.7(3.1)	0.0 0.0
ASIAN/PACIFIC AMERICAN	251	3.4(0.2) < 6%>	52.4(2.8) 284.6(4.8)	47.6(2.8) 297.2(3.6)	0.0 0.0
AMER IND/ALASKAN NATV	38	0.6(0.2) <39%>!	59.4(8.8) 277.3(8.6)	40.6(8.8) 292.4(8.9)	0.0 0.0
PARENTS' EDUCATION LEVEL					
LESS THAN H.S.	503	7.7(0.5) < 8%>	44.2(2.4) 260.4(2.5)	55.8(2.4) 276.9(2.2)	0.0 0.0
GRADUATED H.S.	1441	24.5(1.0) < 4%>	50.7(1.5) 274.9(1.9)	49.3(1.5) 283.7(1.2)	0.0 0.0
SOME EDUC AFTER H.S.	1546	24.7(0.9) < 3%>	45.0(1.3) 288.4(1.8)	55.0(1.3) 294.6(1.3)	0.0 0.0
GRADUATED COLLEGE	2633	40.9(1.4) < 4%>	51.0(1.0) 291.6(1.6)	49.0(1.0) 303.3(1.2)	0.0 0.0
UNKNOWN	129	2.1(0.3) <13%>	50.6(5.7) 246.4(4.6)	49.4(5.7) 267.5(5.3)	0.0 0.0

! INTERPRET WITH CAUTION: SAMPLING ERROR CANNOT BE ESTIMATED ACCURATELY SINCE COEFFICIENT OF VARIATION OF ESTIMATED NUMBER OF STUDENTS EXCEEDS 20%

Table 16-87

Weighted Response Percentages and Proficiency Means (with Standard Errors)
for Main Focused-BIB Reading Samples, by Derived Race/Ethnicity, Grade 12

DERIVED RACE/ETHNICITY	N	WEIGHTED PCT <CV>	WHITE	BLACK	HISPANIC	ASIAN	AMER IND	UNCLASSI	MISSING
-- TOTAL --	6258	100.0(0.0) < 1%>	73.5(0.7) 294.0(1.0)	14.5(0.6) 269.2(1.9)	8.0(0.2) 271.4(2.5)	3.4(0.2) 290.6(3.9)	0.6(0.2) 283.4(7.6)	0.1(0.0) 283.6(10.3)	0.0 0.0
GENDER MALE	3022	49.0(0.8) < 2%>	74.8(1.0) 289.3(1.3)	12.4(0.8) 262.4(2.4)	8.4(0.5) 261.6(2.7)	3.6(0.3) 284.6(4.8)	0.7(0.3) 277.3(8.6)	0.1(0.0) 272.8(9.4)	0.0 0.0
FEMALE	3236	51.0(0.8) < 2%>	72.2(0.8) 298.8(0.9)	16.4(0.7) 274.1(2.1)	7.6(0.4) 281.7(3.1)	3.2(0.3) 297.2(3.6)	0.5(0.2) 292.4(8.9)	0.1(0.0) 297.6(13.5)	0.0 0.0
RACE/ETHNICITY WHITE	4453	73.5(0.7) < 2%>	100.0(0.0) 294.0(1.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
BLACK	844	14.5(0.6) < 4%>	0.0(0.0) 0.0(0.0)	100.0(0.0) 269.2(1.9)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
HISPANIC	666	8.0(C.2) < 3%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 271.4(2.5)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
ASIAN/PACIFIC AMERICAN	251	3.4(0.2) < 6%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 290.6(3.9)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
AMER IND/ALASKAN NATV	38	0.6(0.2) <39%>!	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 283.4(7.6)	0.0(0.0) 0.0(0.0)	0.0 0.0
PARENTS' EDUCATION LEVEL LESS THAN H.S.	503	7.7(0.5) < 8%>	49.4(3.5) 276.4(2.4)	19.4(2.8) 263.3(3.9)	27.3(2.7) 260.7(3.5)	3.2(0.8) 279.0(7.7)	0.7(0.5) 270.8(50.2)	0.0(0.0) 0.0(0.0)	0.0 0.0
GRADUATED H.S.	1441	24.5(1.0) < 4%>	71.6(1.6) 284.3(1.2)	17.7(1.4) 263.8(2.8)	7.1(0.8) 267.7(3.7)	2.7(0.4) 278.9(4.9)	0.8(0.5) 275.1(8.4)	0.1(0.0) 249.5(9.9)	0.0 0.0
SOME EDUC AFTER H.S.	1546	24.7(0.9) < 3%>	74.4(1.2) 295.8(1.2)	15.3(1.0) 276.7(2.7)	7.2(0.6) 285.0(3.5)	2.3(0.5) 286.3(4.9)	0.8(0.3) 284.3(9.4)	0.0(0.0) 292.2(****)	0.0 0.0
GRADUATED COLLEGE	2633	40.9(1.4) < 4%>	80.4(1.1) 301.3(1.0)	10.2(0.7) 272.9(2.8)	4.6(0.5) 278.1(4.1)	4.3(0.3) 301.3(5.4)	0.3(0.2) 299.4(27.0)	0.1(0.1) 293.3(12.4)	0.0 0.0
UNKNOWN	129	2.1(0.3) <13%>	37.1(5.5) 259.1(8.6)	31.4(6.0) 255.7(4.7)	23.5(4.7) 256.0(7.1)	7.4(4.5) 251.7(10.3)	0.5(0.6) 280.8(****)	0.0(0.0) 0.0(0.0)	0.0 0.0

! INTERPRET WITH CAUTION: SAMPLING ERROR CANNOT BE ESTIMATED ACCURATELY SINCE COEFFICIENT OF VARIATION OF ESTIMATED NUMBER OF STUDENTS EXCEEDS 20%

Table 16-88

Weighted Response Percentages and Proficiency Means (with Standard Errors)
for Main Focused-BIB Reading Samples, by Parents' Education Level, Grade 12

PARENTS' EDUCATION LEVEL	N	WEIGHTED PCT <CV>	NOT HS	GRAD HS	POST HS	GRAD COL	UNKNOWN	MISSING
-- TOTAL --	6252	100.0(0.0) < 1%>	7.7(0.5) 269.6(1.9)	24.6(1.0) 279.2(1.0)	24.7(0.9) 291.8(1.1)	41.0(1.4) 297.3(1.1)	2.1(0.3) 256.9(4.4)	0.1 209.6
GENDER MALE	3017	48.9(0.8) < 2%>	6.9(0.6) 260.4(2.5)	25.5(1.1) 274.9(1.9)	22.7(0.9) 288.4(1.8)	42.7(1.5) 291.6(1.6)	2.1(0.4) 246.4(4.6)	0.2 209.4
FEMALE	3235	51.1(0.8) < 2%>	8.4(0.6) 276.9(2.2)	23.7(1.2) 283.7(1.2)	26.6(1.1) 294.6(1.3)	39.3(1.5) 303.3(1.2)	2.0(0.3) 267.5(5.3)	0.0 210.9
RACE/ETHNICITY WHITE	4449	73.5(0.7) < 2%>	5.2(0.5) 276.4(2.4)	24.0(1.0) 284.3(1.2)	25.0(0.9) 295.8(1.2)	44.9(1.7) 301.3(1.0)	1.0(0.2) 259.1(8.6)	0.1 213.3
BLACK	843	14.4(0.6) < 4%>	10.3(1.6) 263.3(3.9)	30.2(2.4) 263.8(2.8)	26.2(1.8) 276.7(2.7)	28.8(1.9) 272.9(2.8)	4.5(1.0) 255.7(4.7)	0.3 206.2
HISPANIC	665	8.0(0.2) < 3%>	26.2(3.0) 260.7(3.5)	21.7(2.6) 267.7(3.7)	22.3(1.8) 285.0(3.5)	23.6(2.4) 278.1(4.1)	6.1(1.3) 256.0(7.1)	0.1 202.9
ASIAN/PACIFIC AMERICAN	251	3.4(0.2) < 6%>	7.1(1.6) 279.0(7.7)	19.3(2.7) 278.9(4.9)	16.7(3.9) 286.3(4.9)	52.3(4.1) 301.3(5.4)	4.5(2.9) 251.7(10.3)	0.0 0.0
AMER IND/ALASKAN NATV	38	0.6(0.2) <39%>!	9.3(6.6) 270.8(50.2)	34.0(8.7) 275.1(8.4)	31.1(8.3) 284.3(9.4)	23.8(9.8) 299.4(27.0)	1.9(1.9) 280.8(****)	0.0 0.0
PARENTS' EDUCATION LEVEL LESS THAN H.S.	503	7.7(0.5) < 8%>	100.0(0.0) 269.6(1.9)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
GRADUATED H.S.	1441	24.6(1.0) < 4%>	0.0(0.0) 0.0(0.0)	100.0(0.0) 279.2(1.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
SOME EDUC AFTER H.S.	1546	24.7(0.9) < 3%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 291.8(1.1)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
GRADUATED COLLEGE	2633	41.0(1.4) < 4%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 297.3(1.1)	0.0(0.0) 0.0(0.0)	0.0 0.0
UNKNOWN	129	2.1(0.3) <13%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 256.9(4.4)	0.0 0.0

! INTERPRET WITH CAUTION: SAMPLING ERROR CANNOT BE ESTIMATED ACCURATELY SINCE COEFFICIENT OF VARIATION OF ESTIMATED NUMBER OF STUDENTS EXCEEDS 20%

Table 16-89

Weighted Response Percentages and Proficiency Means (with Standard Errors)
for Main Focused-BIB Mathematics Samples, by Gender, Grade 4

GENDER OF SUBJECT	N	WEIGHTED PCT <CV>	MALE	FEMALE	MISSING
-- TOTAL --	6467	100.0(0.0) < 0%>	51.7(0.6) 216.7(0.8)	48.3(0.6) 214.9(0.8)	0.0 0.0
GENDER MALE	3349	51.7(0.6) < 1%>	100.0(0.0) 216.7(0.8)	0.0(0.0) 0.0(0.0)	0.0 0.0
FEMALE	3118	48.3(0.6) < 1%>	0.0(0.0) 0.0(0.0)	100.0(0.0) 214.9(0.8)	0.0 0.0
RACE/ETHNICITY WHITE	3794	70.2(0.1) < 0%>	51.2(0.7) 223.5(1.0)	48.8(0.7) 221.7(0.8)	0.0 0.0
BLACK	1120	15.3(0.1) < 1%>	48.3(1.1) 194.4(1.7)	51.7(1.1) 193.7(1.6)	0.0 0.0
HISPANIC	1154	10.7(0.1) < 1%>	56.2(1.9) 201.3(1.8)	43.8(1.9) 199.4(2.0)	0.0 0.0
ASIAN/PACIFIC AMERICAN	228	2.0(0.3) <14%>	58.5(3.7) 227.6(3.0)	41.5(3.7) 229.0(3.7)	0.0 0.0
AMER IND/ALASKAN NATV	158	1.7(0.3) <16%>	65.7(4.7) 213.1(2.9)	34.3(4.7) 205.5(4.3)	0.0 0.0
PARENTS' EDUCATION LEVEL LESS THAN H.S.	298	5.3(0.3) < 6%>	49.5(3.1) 210.2(2.9)	50.5(3.1) 200.2(2.4)	0.0 0.0
GRADUATED H.S.	883	14.9(0.7) < 5%>	52.3(2.3) 210.9(1.4)	47.7(2.3) 211.8(2.0)	0.0 0.0
SOME EDUC AFTER H.S.	501	8.2(0.4) < 5%>	51.6(2.4) 223.5(2.0)	48.4(2.4) 229.6(2.2)	0.0 0.0
GRADUATED COLLEGE	2428	36.0(1.0) < 3%>	53.5(0.9) 223.8(1.2)	46.5(0.9) 221.8(1.0)	0.0 0.0
UNKNOWN	2327	35.1(1.0) < 3%>	49.8(1.2) 210.9(1.1)	50.2(1.2) 208.6(1.0)	0.0 0.0

470 471

Table 16-90

Weighted Response Percentages and Proficiency Means (with Standard Errors)
for Main Focused-BIB Mathematics Samples, by Derived Race/Ethnicity, Grade 4

DERIVED RACE/ETHNICITY	N	WEIGHTED PCT <CV>	WHITE	BLACK	HISPANIC	ASIAN	AMER IND	UNCLASSI	MISSING
-- TOTAL --	6467	100.0(0.0) < 0%>	70.2(0.1) 222.7(0.7)	15.3(0.1) 194.1(1.3)	10.7(0.1) 200.5(1.4)	2.0(0.3) 228.2(2.8)	1.7(0.3) 210.5(2.5)	0.1(0.0) 222.0(8.2)	0.0 0.0
GENDER MALE	3349	51.7(0.6) < 1%>	69.5(0.5) 223.5(1.0)	14.3(0.4) 194.4(1.7)	11.6(0.4) 201.3(1.8)	2.3(0.4) 227.6(3.0)	2.2(0.4) 213.1(2.9)	0.1(0.0) 220.3(14.7)	0.0 0.0
FEMALE	3118	48.3(0.6) < 1%>	70.9(0.5) 221.7(0.8)	16.4(0.4) 193.7(1.6)	9.7(0.4) 199.4(2.0)	1.7(0.2) 229.0(3.7)	1.2(0.3) 205.5(4.3)	0.1(0.0) 224.4(8.0)	0.0 0.0
RACE/ETHNICITY WHITE	3794	70.2(0.1) < 0%>	100.0(0.0) 222.7(0.7)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
BLACK	1120	15.3(0.1) < 1%>	0.0(0.0) 0.0(0.0)	100.0(0.0) 194.1(1.3)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
HISPANIC	1154	10.7(0.1) < 1%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 200.5(1.4)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
ASIAN/PACIFIC AMERICAN	228	2.0(0.3) <14%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 228.2(2.[illegible])	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
AMER IND/ALASKAN NATV	158	1.7(0.3) <16%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 210.5(2.5)	0.0(0.0) 0.0(0.0)	0.0 0.0
PARENTS' EDUCATION LEVEL LESS THAN H.S.	298	5.3(0.3) < 6%>	69.9(2.3) 210.6(2.4)	14.2(2.1) 189.1(5.1)	13.1(1.7) 192.1(3.2)	1.1(0.4) 216.8(8.4)	1.7(0.7) 209.7(8.9)	0.0(0.0) 0.0(0.0)	0.0 0.0
GRADUATED H.S.	883	14.9(0.7) < 5%>	71.5(1.7) 217.7(1.4)	16.6(1.3) 190.2(2.4)	9.1(0.9) 200.7(3.0)	0.7(0.3) 224.4(16.6)	2.0(0.6) 203.6(4.6)	0.1(0.1) 188.7(22.1)	0.0 0.0
SOME EDUC AFTER H.S.	501	8.2(0.4) < 5%>	77.5(1.7) 232.4(1.6)	12.2(1.4) 205.9(4.0)	7.2(0.9) 201.5(4.5)	1.3(0.5) 227.8(9.1)	1.8(0.6) 210.7(8.0)	0.1(0.1) 222.6(****)	0.0 0.0
GRADUATED COLLEGE	2428	36.0(1.0) < 3%>	71.3(0.8) 230.3(1.0)	15.6(0.8) 197.1(1.7)	8.9(0.5) 206.8(2.3)	2.5(0.5) 233.2(3.7)	1.6(0.3) 215.4(3.9)	0.1(0.0) 227.6(4.5)	0.0 0.0
UNKNOWN	2327	35.1(1.0) < 3%>	66.6(0.9) 215.9(0.9)	15.4(0.8) 191.3(2.0)	13.8(0.5) 197.6(2.1)	2.4(0.3) 224.0(3.9)	1.8(0.4) 209.2(4.6)	0.1(0.1) 232.6(14.2)	0.0 0.0

472

473

Table 16-91

Weighted Response Percentages and Proficiency Means (with Standard Errors)
for Main Focused-BIB Mathematics Samples, by Parents' Education Level, Grade 4

PARENTS' EDUCATION LEVEL	N	WEIGHTED PCT <CV>	NGT HS	GRAD HS	POST HS	GRAD COL	UNKNOWN	MISSING
-- TOTAL --	6437	100.0(0.0) < 0%>	5.3(0.3) 205.2(2.0)	15.0(0.7) 211.3(1.2)	8.3(0.4) 226.5(1.6)	36.2(1.0) 222.9(0.9)	35.3(1.0) 209.7(0.9)	0.4 207.7
GENDER MALE	3330	51.6(0.6) < 1%>	5.1(0.4) 210.2(2.9)	15.2(0.9) 210.9(1.4)	8.3(0.6) 223.5(2.0)	37.5(1.1) 223.8(1.2)	34.0(1.1) 210.9(1.1)	0.5 204.8
FEMALE	3107	48.4(0.6) < 1%>	5.6(0.5) 200.2(2.4)	14.7(1.1) 211.8(2.0)	8.3(0.6) 229.6(2.2)	34.8(1.2) 221.8(1.0)	33.6(1.3) 208.6(1.0)	0.3 212.3
RACE/ETHNICITY WHITE	3771	70.1(0.1) < 1%>	5.3(0.4) 210.6(2.4)	15.2(0.9) 217.7(1.4)	9.1(0.5) 232.4(1.6)	36.8(1.1) 230.3(1.0)	33.5(1.1) 215.9(0.9)	0.5 212.7
BLACK	1118	15.3(0.1) < 1%>	5.0(0.8) 189.1(5.1)	16.2(1.4) 190.2(2.4)	6.6(0.8) 205.9(4.0)	36.8(2.0) 197.1(1.7)	35.4(2.2) 191.3(2.0)	0.2 192.2
HISPANIC	1149	10.7(0.1) < 1%>	6.5(0.8) 192.1(3.2)	12.7(1.2) 200.7(3.0)	5.5(0.7) 201.5(4.5)	29.9(1.6) 206.8(2.3)	45.4(1.9) 197.6(2.1)	0.4 178.7
ASIAN/PACIFIC AMERICAN	228	2.0(0.3) <14%>	2.8(1.4) 216.8(8.4)	5.0(1.7) 224.4(16.6)	5.2(1.6) 227.8(9.1)	45.0(3.9) 233.2(3.7)	42.0(4.5) 224.0(3.9)	0.0 0.0
AMER IND/ALASKAN NATV	158	1.8(0.3) <16%>	5.3(2.0) 209.7(8.9)	17.0(3.2) 203.6(4.6)	8.3(3.0) 210.7(8.0)	34.0(3.8) 215.4(3.9)	35.3(4.7) 209.2(4.6)	0.0 0.0
PARENTS' EDUCATION LEVEL LESS THAN H.S.	298	5.3(0.3) < 6%>	100.0(0.0) 205.2(2.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
GRADUATED H.S.	883	15.0(0.7) < 5%>	0.0(0.0) 0.0(0.0)	100.0(0.0) 211.3(1.2)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
SOME EDUC AFTER H.S.	501	8.3(0.4) < 5%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 226.5(1.6)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
GRADUATED COLLEGE	2428	36.2(1.0) < 3%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 222.9(0.9)	0.0(0.0) 0.0(0.0)	0.0 0.0
UNKNOWN	2327	35.3(1.0) < 3%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 209.7(0.9)	0.0 0.0

474

475

Table 16-92

Weighted Response Percentages and Proficiency Means (with Standard Errors) for Main Focused-BIB Mathematics Samples, by Gender, Grade 8

GENDER OF SUBJECT	N	WEIGHTED PCT <CV>	MALE	FEMALE	MISSING
-- TOTAL --	6473	100.0(0.0) < 0%>	50.0(0.7) 265.5(1.3)	50.0(0.7) 264.4(1.1)	0.0 0.0
GENDER MALE	3218	50.0(0.7) < 1%>	100.0(0.0) 265.5(1.3)	0.0(0.0) 0.0(0.0)	0.0 0.0
FEMALE	3255	50.0(0.7) < 1%>	0.0(0.0) 0.0(0.0)	100.0(0.0) 264.4(1.1)	0.0 0.0
RACE/ETHNICITY WHITE	4220	70.6(0.2) < 0%>	50.0(0.9) 272.9(1.5)	50.0(0.9) 271.4(1.2)	0.0 0.0
BLACK	944	15.1(0.1) < 1%>	46.2(1.7) 241.7(1.9)	53.8(1.7) 239.9(1.9)	0.0 0.0
HISPANIC	920	10.0(0.1) < 2%>	54.3(1.9) 247.4(1.9)	45.7(1.9) 248.4(1.8)	0.0 0.0
ASIAN/PACIFIC AMERICAN	267	2.7(0.4) <13%>	47.7(3.1) 283.3(5.7)	52.3(3.1) 286.2(3.7)	0.0 0.0
AMER IND/ALASKAN NATV	100	1.4(0.4) <28%>!	67.0(5.5) 248.8(6.4)	33.0(5.5) 246.1(9.6)	0.0 0.0
PARENTS' EDUCATION LEVEL LESS THAN H.S.	542	8.9(0.6) < 7%>	42.4(2.3) 247.0(2.1)	57.6(2.3) 245.0(1.7)	0.0 0.0
GRADUATED H.S.	1516	24.5(0.9) < 4%>	49.6(1.5) 255.8(1.4)	50.4(1.5) 255.2(1.1)	0.0 0.0
SOME EDUC AFTER H.S.	1149	17.9(0.6) < 3%>	45.4(1.7) 270.4(1.7)	54.6(1.7) 270.0(1.2)	0.0 0.0
GRADUATED COLLEGE	2690	40.3(1.5) < 4%>	52.3(1.2) 277.4(1.7)	47.7(1.2) 277.0(1.6)	0.0 0.0
UNKNOWN	557	8.1(0.4) < 5%>	58.3(2.2) 244.2(2.6)	41.7(2.2) 240.4(2.1)	0.0 0.0

! INTERPRET WITH CAUTION: SAMPLING ERROR CANNOT BE ESTIMATED ACCURATELY SINCE COEFFICIENT OF VARIATION OF ESTIMATED NUMBER OF STUDENTS EXCEEDS 20%

Table 16-93

Weighted Response Percentages and Proficiency Means (with Standard Errors)
for Main Focused-BIB Mathematics Samples, by Derived Race/Ethnicity, Grade 8

DERIVED RACE/ETHNICITY	N	WEIGHTED PCT <CV>	WHITE	BLACK	HISPANIC	ASIAN	AMER IND	UNCLASSI	MISSING
-- TOTAL --	6473	100.0(0.0) < 0%>	70.6(0.2) 272.1(1.2)	15.1(0.1) 240.8(1.6)	10.0(0.1) 247.9(1.6)	2.7(0.4) 284.8(4.1)	1.4(0.4) 247.9(3.4)	0.2(0.1) 270.0(7.1)	0.0 0.0
GENDER									
MALE	3218	50.0(0.7) < 1%>	70.6(0.6) 272.9(1.5)	13.9(0.5) 241.7(1.9)	10.9(0.4) 247.4(1.9)	2.5(0.4) 283.3(5.7)	1.9(0.6) 248.8(6.4)	0.2(0.1) 252.5(12.1)	0.0 0.0
FEMALE	3255	50.0(0.7) < 1%>	70.6(0.7) 271.4(1.2)	16.2(0.5) 239.9(1.9)	9.2(0.4) 248.4(1.8)	2.8(0.3) 286.2(3.7)	0.9(0.3) 246.1(9.6)	0.3(0.1) 283.3(6.1)	0.0 0.0
RACE/ETHNICITY									
WHITE	4220	70.6(0.2) < 0%>	100.0(0.0) 272.1(1.2)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
BLACK	944	15.1(0.1) < 1%>	0.0(0.0) 0.0(0.0)	100.0(0.0) 240.8(1.6)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
HISPANIC	920	10.0(0.1) < 2%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 247.9(1.6)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
ASIAN/PACIFIC AMERICAN	267	2.7(0.4) <13%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 284.8(4.1)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
AMER IND/ALASKAN NATV	100	1.4(0.4) <28%>!	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 247.9(3.4)	0.0(0.0) 0.0(0.0)	0.0 0.0
PARENTS' EDUCATION LEVEL									
LESS THAN H.S.	542	8.9(0.6) < 7%>	58.3(3.4) 249.4(1.4)	17.4(2.4) 234.6(4.3)	19.1(2.5) 244.3(4.1)	1.6(0.5) 272.5(7.2)	3.5(1.2) 238.5(8.4)	0.1(0.1) 261.3(****)	0.0 0.0
GRADUATED H.S.	1516	24.5(0.9) < 4%>	72.0(1.4) 261.6(1.0)	16.2(1.2) 233.9(2.0)	9.2(0.6) 245.5(2.0)	1.2(0.2) 269.1(7.9)	1.2(0.3) 241.7(6.9)	0.2(0.1) 268.4(15.6)	0.0 0.0
SOME EDUC AFTER H.S.	1149	17.9(0.6) < 3%>	75.2(1.3) 275.8(1.1)	14.1(1.0) 247.5(2.4)	8.6(0.6) 257.9(2.8)	0.8(0.3) 280.7(17.[illegible])	1.1(0.5) 268.7(9.3)	0.1(0.1) 277.4(30.3)	0.0 0.0
GRADUATED COLLEGE	2690	40.3(1.5) < 4%>	74.6(1.0) 283.4(1.5)	13.3(0.9) 249.4(?.5)	6.6(0.5) 256.6(2.3)	4.2(0.7) 294.1(5.0)	1.0(0.4) 256.6(8.5)	0.3(0.1) 271.0(11.1)	0.0 0.0
UNKNOWN	557	8.1(0.4) < 5%>	50.3(2.9) 251.7(2.3)	19.9(?.1) 225.9([illegible])	22.1(1.9) 233.3(2.8)	4.7(1.1) 262.7(6.2)	2.8(1.0) 235.8(8.1)	0.3(0.2) 261.5(23.0)	0.0 0.0

! INTERPRET WITH CAUTION: SAMPLING ERROR CANNOT BE ESTIMATED ACCURATELY SINCE COEFFICIENT OF VARIATION OF ESTIMATED NUMBER OF STUDENTS EXCEEDS 20%

Table 16-94

Weighted Response Percentages and Proficiency Means (with Sta   rd Errors)
for Main Focused-BIB Mathematics Samples, by Parents' Education Level, Grade 8

PARENTS' EDUCATION LEVEL	N	WEIGHTED PCT <CV>	NOT HS	GRAD HS	POST HS	GRAD COL	UNKNOWN	MISSING
-- TOTAL --	6454	100.0(0.0) < 0%>	8.9(0.6) 245.8(1.4)	24.6(0.9) 255.5(0.9)	18.0(0.6) 270.2(1.0)	40.4(1.5) 277.2(1.5)	8.1(0.4) 242.6(1.9)	0.3 246.1
GENDER								
MALE	3206	50.0(0.7) < 1%>	7.5(0.7) 247.0(2.1)	24.4(1.1) 255.8(1.4)	16.3(0.8) 270.4(1.7)	42.3(1.6) 277.4(1.7)	9.5(0.6) 244.2(2.6)	0.4 259.8
FEMALE	3248	50.0(0.7) < 1%>	10.2(0.8) 245.0(1.7)	24.8(1.2) 255.2(1.1)	19.6(0.9) 270.0(1.2)	38.6(1.9) 277.0(1.6)	6.8(0.5) 240.4(2.1)	0.2 221.9
RACE/ETHNICITY								
WHITE	4210	70.6(0.2) < 0%>	7.3(0.7) 249.4(1.4)	25.1(1.2) 261.6(1.0)	19.1(0.7) 275.8(1.1)	42.7(1.8) 283.4(1.5)	5.8(0.5) 251.7(2.3)	0.3 271.4
BLACK	940	15.1(0.1) < 1%>	10.3(1.6) 234.6(4.3)	26.4(2.2) 233.9(2.0)	16.8(1.4) 247.5(2.4)	35.8(2.3) 249.4(2.5)	10.8(1.3) 225.9(3.4)	0.4 184.2
HISPANIC	915	10.0(0.1) < 2%>	17.0(2.5) 244.3(4.1)	22.7(1.3) 245.5(2.0)	15.5(1.2) 257.9(2.8)	26.7(2.2) 256.6(2.3)	18.0(1.4) 233.3(2.8)	0.5 218.4
ASIAN/PACIFIC AMERICAN	267	2.7(0.4) <13%>	5.3(1.8) 272.5(7.2)	11.5(2.0) 269.1(7.9)	5.5(1.6) 280.7(17.8)	63.4(4.7) 294.1(5.0)	14.4(3.5) 262.7(6.2)	0.0 0.0
AMER IND/ALASKAN NATV	100	1.4(0.4) <28%>!	21.8(4.2) 238.5(8.4)	21.0(7.5) 241.7(6.9)	14.0(3.5) 268.7(9.3)	27.3(5.1) 256.6(8.5)	15.9(3.3) 235.8(8.1)	0.0 0.0
PARENTS' EDUCATION LEVEL								
LESS THAN H.S.	542	8.9(0.6) < 7%>	100.0(0.0) 245.8(1.4)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
GRADUATED H.S.	1516	24.6(0.9) < 4%>	0.0(0.0) 0.0(0.0)	100.0(0.0) 255.5(0.9)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
SOME EDUC AFTER H.S.	1149	18.0(0.6) < 3%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 270.2(1.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
GRADUATED COLLEGE	2690	40.4(1.5) < 4%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 277.2(1.5)	0.0(0.0) 0.0(0.0)	0.0 0.0
UNKNOWN	557	8.1(0.4) < 5%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 242.6(1.9)	0.0 0.0

! INTERPRET WITH CAUTION: SAMPLING ERROR CANNOT BE ESTIMATED ACCURATELY SINCE COEFFICIENT OF VARIATION OF ESTIMATED NUMBER OF STUDENTS EXCEEDS 20%

Table 16-95

Weighted Response Percentages and Proficiency Means (with Standard Errors)
for Main Focused-BIB Mathematics Samples, by Gender, Grade 12

GENDER OF SUBJECT	N	WEIGHTED PCT <CV>	MALE	FEMALE	MISSING
-- TOTAL --	6311	100.0(0.0) < 1%>	48.8(0.7) 297.7(1.3)	51.2(0.7) 293.1(1.1)	0.0 0.0
GENDER					
MALE	3065	48.8(0.7) < 2%>	100.0(0.0) 297.7(1.3)	0.0(0.0) 0.0(0.0)	0.0 0.0
FEMALE	3246	51.2(0.7) < 1%>	0.0(0.0) 0.0(0.0)	100.0(0.0) 293.1(1.1)	0.0 0.0
RACE/ETHNICITY					
WHITE	4492	73.9(0.4) < 1%>	48.9(0.8) 303.0(1.5)	51.1(0.8) 299.3(1.2)	0.0 0.0
BLACK	913	14.0(0.3) < 3%>	47.5(1.8) 273.1(1.7)	52.5(1.8) 267.6(1.4)	0.0 0.0
HISPANIC	611	7.9(0.3) < 4%>	49.6(2.8) 281.7(2.4)	50.4(2.8) 273.6(3.3)	0.0 0.0
ASIAN/PACIFIC AMERICAN	245	3.4(0.2) < 6%>	48.5(3.0) 316.3(4.6)	51.5(3.0) 313.9(4.6)	0.0 0.0
AMER IND/ALASKAN NATV	45	0.8(0.3) <37%>!	53.7(8.4) 298.7(5.9)	46.3(8.4) 280.8(8.0)	0.0 0.0
PARENTS' EDUCATION LEVEL					
LESS THAN H.S.	511	7.9(0.6) < 8%>	43.7(2.3) 276.0(2.3)	56.3(2.3) 269.6(1.8)	0.0 0.0
GRADUATED H.S.	1420	23.5(1.0) < 4%>	49.0(1.5) 284.9(1.6)	51.0(1.5) 279.7(1.4)	0.0 0.0
SOME EDUC AFTER H.S.	1591	25.3(0.8) < 3%>	45.8(1.5) 298.5(1.3)	54.2(1.5) 295.6(1.0)	0.0 0.0
GRADUATED COLLEGE	2644	40.9(1.2) < 3%>	50.6(0.9) 310.0(1.7)	49.4(0.9) 306.0(1.3)	0.0 0.0
UNKNOWN	126	2.1(0.2) <10%>	64.8(4.3) 270.2(3.3)	35.2(4.3) 264.1(7.3)	0.0 0.0

! INTERPRET WITH CAUTION: SAMPLING ERROR CANNOT BE ESTIMATED ACCURATELY SINCE COEFFICIENT OF VARIATION OF ESTIMATED NUMBER OF STUDENTS EXCEEDS 20%

Table 16-96

Weighted Response Percentages and Proficiency Means (with Standard Errors)
for Main Focused-BIB Mathematics Samples, by Derived Race/Ethnicity, Grade 12

DERIVED RACE/ETHNICITY	N	WEIGHTED PCT <CV>	WHITE	BLACK	HISPANIC	ASIAN	AMER IND	UNCLASSI	MISSING
-- TOTAL --	6311	100.0(0.0) < 1%>	73.9(0.4)	14.0(0.3)	7.9(0.3)	3.4(0.2)	0.8(0.3)	0.0(0.0)	0.0
			301.1(1.2)	270.2(1.3)	277.6(2.4)	315.0(4.0)	290.4(5.4)	303.3(8.9)	0.0
GENDER									
MALE	3065	48.8(0.7) < 2%>	74.0(0.7)	13.7(0.5)	8.0(0.5)	3.4(0.3)	0.9(0.3)	0.0(0.0)	0.0
			303.0(1.5)	273.1(1.7)	281.7(2.4)	316.3(4.6)	298.7(5.9)	330.8(****)	0.0
FEMALE	3246	51.2(0.7) < 1%>	73.7(0.7)	14.4(0.5)	7.8(0.5)	3.4(0.2)	0.7(0.3)	0.1(0.0)	0.0
			299.3(1.2)	267.6(1.4)	273.6(3.3)	313.9(4.6)	280.8(8.0)	297.4(3.7)	0.0
RACE/ETHNICITY									
WHITE	4492	73.9(0.4) < 1%>	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
			301.1(1.2)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
BLACK	913	14.0(0.3) < 3%>	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
			0.0(0.0)	270.2(1.3)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
HISPANIC	611	7.9(0.3) < 4%>	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
			0.0(0.0)	0.0(0.0)	277.6(2.4)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
ASIAN/PACIFIC AMERICAN	245	3.4(0.2) < 6%>	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
			0.0(0.0)	0.0(0.0)	0.0(0.0)	315.0(4.0)	0.0(0.0)	0.0(0.0)	0.0
AMER IND/ALASKAN NATV	45	0.8(0.3) <37%>!	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0
			0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	290.4(5.4)	0.0(0.0)	0.0
PARENTS' EDUCATION LEVEL									
LESS THAN H.S.	511	7.9(0.6) < 8%>	50.6(3.3)	18.9(3.0)	25.0(3.0)	3.6(1.0)	1.6(0.8)	0.2(0.2)	0.0
			276.4(2.6)	261.8(3.0)	266.7(2.7)	306.8(12.1)	279.8(5.9)	292.5(14.7)	0.0
GRADUATED H.S.	1420	23.5(1.0) < 4%>	73.0(1.5)	16.3(1.1)	8.3(1.1)	1.7(0.5)	0.6(0.3)	0.0(0.0)	0.0
			287.1(1.4)	261.4(1.9)	276.7(4.5)	303.3(7.6)	269.9(14.2)	0.0(0.0)	0.0
SOME EDUC AFTER H.S.	1591	25.3(0.8) < 3%>	77.8(1.1)	14.9(0.9)	4.5(0.5)	2.4(0.6)	0.4(0.3)	0.0(0.0)	0.0
			301.3(1.0)	274.9(2.2)	287.6(3.5)	308.3(7.6)	300.3(7.7)	0.0(0.0)	0.0
GRADUATED COLLEGE	2644	40.9(1.2) < 3%>	77.8(0.9)	10.9(0.7)	5.5(0.5)	4.8(0.4)	0.9(0.4)	0.0(0.0)	0.0
			312.7(1.4)	278.8(2.0)	289.2(3.5)	321.5(3.9)	302.6(4.3)	313.4(8.8)	0.0
UNKNOWN	126	2.1(0.2) <10%>	47.7(5.0)	18.4(3.0)	25.3(3.3)	7.4(3.3)	1.2(1.2)	0.0(0.0)	0.0
			277.2(3.7)	252.8(5.2)	252.1(8.1)	305.0(18.4)	247.4(8.0)	0.0(0.0)	0.0

! INTERPRET WITH CAUTION: SAMPLING ERROR CANNOT BE ESTIMATED ACCURATELY SINCE COEFFICIENT OF VARIATION OF ESTIMATED NUMBER OF STUDENTS EXCEEDS 20%

485

484

Table 16-97

Weighted Response Percentages and Proficiency Means (with Standard Errors)
for Main Focused-BIB Mathematics Samples, by Parents' Education Level, Grade 12

PARENTS' EDUCATION	N	WEIGHTED PCT <CV>	NOT HS	GRAD HS	POST HS	GRAD COL	UNKNOWN	MISSING
-- TOTAL --	6292	100.0(0.0) < 1%>	7.9(0.6) 272.4(1.4)	23.6(1.0) 282.2(1.2)	25.4(0.8) 296.9(0.9)	41.0(1.2) 308.1(1.3)	2.1(0.2) 268.0(3.5)	0.3 253.1
GENDER								
MALE	3056	48.8(0.7) < 2%>	7.1(0.6) 276.0(2.3)	23.7(1.3) 284.9(1.6)	23.8(0.9) 298.5(1.3)	42.5(1.4) 310.0(1.7)	2.8(0.3) 270.2(3.3)	0.3 253.8
FEMALE	3236	51.2(0.7) < 1%>	8.7(0.7) 269.6(1.8)	23.5(1.0) 279.7(1.4)	26.8(1.1) 295.6(1.0)	39.5(1.3) 306.0(1.3)	1.4(0.2) 264.1(7.3)	0.3 252.5
RACE/ETHNICITY								
WHITE	4482	73.9(0.4) < 1%>	5.4(0.5) 276.4(2.6)	23.3(1.2) 287.1(1.4)	26.7(0.9) 301.3(1.0)	43.2(1.4) 312.7(1.4)	1.4(0.2) 277.2(3.7)	0.2 258.5
BLACK	907	14.0(0.3) < 3%>	10.7(1.7) 261.8(3.0)	27.5(1.6) 261.4(1.9)	27.1(1.6) 274.9(2.2)	31.9(2.0) 278.8(2.0)	2.8(0.5) 252.8(5.2)	0.5 233.1
HISPANIC	608	7.9(0.3) < 4%>	25.2(3.4) 266.7(2.7)	24.9(3.7) 276.7(4.5)	14.3(1.8) 287.6(3.5)	28.7(2.3) 289.2(3.5)	6.8(1.0) 252.1(8.1)	0.4 270.8
ASIAN/PACIFIC AMERICAN	245	3.4(0.2) < 6%>	8.4(1.9) 306.8(12.1)	11.6(3.3) 303.3(7.6)	17.8(4.4) 308.3(7.6)	57.7(3.6) 321.5(3.9)	4.5(2.3) 305.0(18.4)	0.0 0.0
AMER IND/ALASKAN NATV	45	0.8(0.3) <37%>!	16.7(6.2) 279.8(5.9)	19.2(7.5) 269.9(14.2)	14.0(5.1) 300.3(7.7)	46.9(10.1) 302.6(4.3)	3.2(3.1) 247.4(8.0)	0.0 0.0
PARENTS' EDUCATION LEVEL								
LESS THAN H.S.	511	7.9(0.6) < 8%>	100.0(0.0) 272.4(1.4)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
GRADUATED H.S.	1420	23.6(1.0) < 4%>	0.0(0.0) 0.0(0.0)	100.0(0.0) 282.2(1.2)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
SOME EDUC AFTER H.S.	1591	25.4(0.8) < 3%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 296.9(0.9)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
GRADUATED COLLEGE	2644	41.0(1.2) < 3%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 308.1(1.3)	0.0(0.0) 0.0(0.0)	0.0 0.0
UNKNOWN	126	2.[illegible] 0.2) <10%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 268.0(3.5)	0.0 0.0

! INTERPRET WITH CAUTION: SAMPLING ERROR CANNOT BE ESTIMATED ACCURATELY SINCE COEFFICIENT OF VARIATION OF ESTIMATED NUMBER OF STUDENTS EXCEEDS 20%

Table 16-98

Weighted Response Percentages and Proficiency Means (with Standard Errors)
for Main Focused-BIB Science Samples, by Gender, Grade 4

GENDER OF SUBJECT	N	WEIGHTED PCT <CV>	MALE	FEMALE	MISSING
-- TOTAL --	6314	100.0(0.0) < 1%>	51.2(0.7)	48.8(0.7)	0.0
			233.8(1.1)	231.7(1.0)	0.0
GENDER					
MALE	3220	51.2(0.7) < 2%>	100.0(0.0)	0.0(0.0)	0.0
			233.8(1.1)	0.0(0.0)	0.0
FEMALE	3094	48.8(0.7) < 2%>	0.0(0.0)	100.0(0.0)	0.0
			0.0(0.0)	231.7(1.0)	0.0
RACE/ETHNICITY					
WHITE	3715	70.2(0.5) < 1%>	51.5(1.0)	48.5(1.0)	0.0
			243.2(1.3)	240.9(1.1)	0.0
BLACK	1075	15.2(0.4) < 2%>	48.7(1.6)	51.3(1.6)	0.0
			204.9(1.8)	205.8(1.8)	0.0
HISPANIC	1161	11.0(0.3) < 3%>	51.9(1.6)	48.1(1.6)	0.0
			213.0(1.6)	211.0(1.9)	0.0
ASIAN/PACIFIC AMERICAN	199	1.9(0.3) <16%>	51.0(4.0)	49.0(4.0)	0.0
			231.6(3.2)	234.9(4.4)	0.0
AMER IND/ALASKAN NATV	157	1.6(0.3) <18%>	57.5(4.2)	42.5(4.2)	0.0
			227.1(3.4)	224.8(4.0)	0.0
PARENTAL EDUCATION					
LESS THAN H.S.	307	5.1(0.4) < 8%>	50.3(3.4)	49.7(3.4)	0.0
			222.2(3.4)	220.6(2.8)	0.0
GRADUATED H.S.	928	15.7(0.7) < 4%>	51.3(1.7)	48.7(1.7)	0.0
			228.2(1.8)	223.1(1.9)	0.0
SOME EDUC AFTER H.S.	524	8.6(0.4) < 5%>	50.8(2.7)	49.2(2.7)	0.0
			243.7(2.5)	240.1(2.6)	0.0
GRADUATED COLLEGE	2290	35.2(1.1) < 3%>	54.0(1.3)	46.0(1.3)	0.0
			243.0(1.5)	242.2(1.4)	0.0
UNKNOWN	2224	34.8(0.8) < 2%>	48.3(1.1)	51.7(1.1)	0.0
			225.8(1.2)	225.8(1.1)	0.0

Table 16-99

Weighted Response Percentages and Proficiency Means (with Standard Errors)
for Main Focused-BIB Science Samples, by Derived Race/Ethnicity, Grade 4

DERIVED RACE/ETHNICITY	N	WEIGHTED PCT <CV>	WHITE	BLACK	HISPANIC	ASIAN	AMER IND	UNCLASSIFIED	MISSING
-- TOTAL --	6314	100.0(0.0) < 1%>	70.2(0.5) 242.1(1.0)	15.2(0.4) 205.4(1.5)	11.0(0.3) 212.0(1.5)	1.9(0.3) 233.2(3.0)	1.6(0.3) 226.1(2.7)	0.1(0.0) 223.0(24.3)	0.0 0.0
GENDER									
MALE	3220	51.2(0.7) < 2%>	70.7(0.7) 243.2(1.3)	14.4(0.5) 204.9(1.8)	11.1(0.5) 213.0(1.6)	1.9(0.3) 231.6(3.2)	1.8(0.3) 227.1(3.4)	0.1(0.0) 250.3(3.8)	0.0 0.0
FEMALE	3094	48.8(0.7) < 2%>	69.7(0.7) 240.9(1.1)	16.6(0.7) 205.8(1.8)	10.8(0.5) 211.0(1.9)	1.9(0.3) 234.9(4.4)	1.4(0.3) 224.8(4.0)	0.2(0.1) 212.7(32.7)	0.0 0.0
RACE/ETHNICITY									
WHITE	3715	70.2(0.5) < 1%>	100.0(0.0) 242.1(1.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
BLACK	1075	15.2(0.4) < 2%>	0.0(0.0) 0.0(0.0)	100.0(0.0) 205.4(1.5)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
HISPANIC	1161	11.0(0.3) < 3%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 212.0(1.5)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
ASIAN/PACIFIC AMERICAN	199	1.9(0.3) <16%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 233.2(3.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
AMER IND/ALASKAN NATV	157	1.6(0.3) <18%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 226.1(2.7)	0.0(0.0) 0.0(0.0)	0.0 0.0
PARENTAL EDUCATION									
LESS THAN H.S.	307	5.1(0.4) < 8%>	67.8(3.4) 228.4(2.7)	12.5(2.3) 206.3(5.3)	16.4(2.3) 204.8(3.3)	1.5(0.7) 226.4(11.4)	1.8(0.9) 209.8(7.2)	0.0(0.0) 0.0(0.0)	0.0 0.0
GRADUATED H.S.	928	15.7(0.7) < 4%>	69.8(1.7) 234.5(1.5)	16.8(1.2) 201.9(2.6)	10.8(0.9) 208.2(3.1)	0.7(0.2) 219.9(7.0)	1.7(0.5) 219.4(6.8)	0.2(0.2) 187.3(****)	0.0 0.0
SOME EDUC AFTER H.S.	524	8.6(0.4) < 5%>	77.0(1.6) 249.6(1.8)	10.9(1.4) 206.8(4.5)	9.3(1.1) 220.7(4.1)	1.5(0.5) 243.3(8.1)	1.3(0.4) 231.2(7.5)	0.0(0.0) 0.0(0.0)	0.0 0.0
GRADUATED COLLEGE	2290	35.2(1.1) < 3%>	71.2(1.1) 252.7(1.2)	16.6(0.8) 210.7(2.3)	8.6(0.5) 221.8(2.2)	2.0(0.4) 246.0(5.1)	1.4(0.3) 234.4(5.2)	0.1(0.1) 246.4(6.3)	0.0 0.0
UNKNOWN	2224	34.8(0.8) < 2%>	68.0(0.8) 234.7(1.0)	14.3(0.8) 201.1(2.0)	13.2(0.6) 207.0(2.1)	2.5(0.4) 224.2(3.2)	1.9(0.4) 223.9(4.0)	0.1(0.1) 233.6(12.6)	0.0 0.0

490

491

Table 16-100

Weighted Response Percentages and Proficiency Means (with Standard Errors)
for Main Focused-BIB Science Samples, by Parents' Education Level, Grade 4

PARENTS' EDUCATION LEVEL	N	WEIGHTED PCT <CV>	NOT HS	GRAD HS	POST HS	GRAD COL	UNKNOWN	MISSING
-- TOTAL --	6273	100.0(0.0) < 1%>	5.2(0.4) 221.4(2.2)	15.8(0.7) 225.7(1.4)	8.7(0.4) 241.9(1.8)	35.4(1.1) 242.7(1.2)	35.0(0.8) 225.8(0.8)	0.6 207.1
GENDER								
MALE	3193	51.1(0.8) < 2%>	5.1(0.5) 222.2(3.4)	15.8(0.8) 228.2(1.8)	8.6(0.6) 243.7(2.5)	37.4(1.2) 243.0(1.5)	33.1(1.1) 225.8(1.2)	0.7 211.5
FEMALE	3080	48.9(0.8) < 2%>	5.3(0.6) 220.6(2.8)	15.7(0.9) 223.1(1.9)	8.7(0.6) 240.1(2.6)	33.3(1.4) 242.2(1.4)	37.0(1.0) 225.8(1.1)	0.5 200.2
RACE/ETHNICITY								
WHITE	3690	70.2(0.5) < 1%>	5.0(0.5) 228.4(2.7)	15.7(0.8) 234.5(1.5)	9.5(0.5) 249.6(1.8)	35.9(1.3) 252.7(1.2)	33.9(0.9) 234.7(1.0)	0.6 214.7
BLACK	1065	15.1(0.4) < 2%>	4.3(0.9) 206.3(5.3)	17.5(1.2) 201.9(2.6)	6.3(0.8) 206.8(4.5)	38.9(1.7) 210.7(2.3)	33.1(1.6) 201.1(2.0)	0.9 187.2
HISPANIC	1156	11.0(0.3) < 3%>	7.7(1.1) 204.8(3.3)	15.4(1.2) 208.2(3.1)	7.3(0.8) 220.7(4.1)	27.6(2.1) 221.8(2.2)	42.0(1.9) 207.0(2.1)	0.3 190.7
ASIAN/PACIFIC AMERICAN	198	1.9(0.3) <16%>	4.1(1.7) 226.4(11.4)	6.1(1.6) 219.9(7.0)	6.7(1.5) 243.3(8.1)	37.1(3.3) 246.0(5.1)	46.0(3.6) 224.2(3.2)	0.4 181.1
AMER IND/ALASKAN NATV	157	1.6(0.3) <18%>	5.8(2.6) 209.8(7.2)	15.9(3.3) 219.4(6.8)	6.7(1.8) 231.2(7.5)	30.9(3.5) 234.4(5.2)	40.8(4.5) 223.9(4.0)	0.0 0.0
PARENTAL EDUCATION								
LESS THAN H.S.	307	5.2(0.4) < 8%>	100.0(0.0) 221.4(2.2)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
GRADUATED H.S.	928	15.8(0.7) < 4%>	0.0(0.0) 0.0(0.0)	100.0(0.0) 225.7(1.4)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
SOME EDUC AFTER H.S.	524	8.7(0.4) < 5%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 241.9(1.8)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
GRADUATED COLLEGE	2290	35.4(1.1) < 3%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 242.7(1.2)	0.0(0.0) 0.0(0.0)	0.0 0.0
UNKNOWN	2224	35.0(0.8) < 2%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 225.8(0.8)	0.0 0.0

Table 16-101

Weighted Response Percentages and Proficiency Means (with Standard Errors)
for Main Focused-BIB Science Samples, by Gender, Grade 8

GENDER OF SUBJECT	N	WEIGHTED PCT <CV>	MALE	FEMALE	MISSING
-- TOTAL --	6531	100.0(0.0) < 1%>	50.1(0.8) 265.1(1.6)	49.9(0.8) 261.0(1.2)	0.0 0.0
GENDER					
MALE	3228	50.1(0.8) < 2%>	100.0(0.0) 265.1(1.6)	0.0(0.0) 0.0(0.0)	0.0 0.0
FEMALE	3303	49.9(0.8) < 2%>	0.0(0.0) 0.0(0.0)	100.0(0.0) 261.0(1.2)	0.0 0.0
RACE/ETHNICITY					
WHITE	4223	70.8(0.4) < 1%>	50.9(1.0) 274.4(1.8)	49.1(1.0) 271.3(1.4)	0.0 0.0
BLACK	917	14.8(0.4) < 3%>	45.1(1.7) 231.9(2.9)	54.9(1.7) 230.3(2.1)	0.0 0.0
HISPANIC	1000	10.1(0.3) < 3%>	51.9(1.6) 243.1(3.0)	48.1(1.6) 239.1(2.5)	0.0 0.0
ASIAN/PACIFIC AMERICAN	285	2.7(0.4) <14%>	52.9(3.9) 272.5(4.7)	47.1(3.8) 268.3(4.5)	0.0 0.0
AMER IND/ALASKAN NATV	95	1.4(0.5) <36%>!	42.0(5.4) 255.2(10.3)	58.0(5.4) 249.5(7.8)	0.0 0.0
PARENTAL EDUCATION					
LESS THAN H.S.	584	8.8(0.6) < 7%>	40.4(2.3) 244.8(3.3)	59.6(2.3) 238.7(2.8)	0.0 0.0
GRADUATED H.S.	1540	24.8(0.8) < 3%>	50.1(1.4) 255.9(1.8)	49.9(1.4) 252.3(1.5)	0.0 0.0
SOME EDUC AFTER H.S.	1195	18.7(0.8) < 4%>	46.0(1.4) 270.3(2.3)	54.0(1.4) 266.7(1.5)	0.0 0.0
GRADUATED COLLEGE	2642	39.6(1.6) < 4%>	52.6(1.2) 277.4(2.0)	47.4(1.2) 275.3(1.9)	0.0 0.0
UNKNOWN	559	7.9(0.4) < 5%>	57.2(2.7) 240.6(3.0)	42.8(2.7) 231.0(2.7)	0.0 0.0

! INTERPRET WITH CAUTION: SAMPLING ERROR CANNOT BE ESTIMATED ACCURATELY SINCE COEFFICIENT OF VARIATION OF ESTIMATED NUMBER OF STUDENTS EXCEEDS 20%

Table 16-102

Weighted Response Percentages and Proficiency Means (with Standard Errors)
for Main Focused-BIB Science Samples, by Derived Race/Ethnicity, Grade 8

DERIVED RACE/ETHNICITY	N	WEIGHTED PCT <CV>	WHITE	BLACK	HISPANIC	ASIAN	AMER IND	UNCLASSIFIED	MISSING
-- TOTAL --	6531	100.0(0.0) < 1%>	70.8(0.4)	14.8(0.4)	10.1(0.3)	2.7(0.4)	1.4(0.5)	0.1(0.0)	0.0
			272.9(1.4)	231.0(2.2)	241.2(2.1)	270.5(4.0)	251.9(8.5)	255.1(12.8)	0.0
GENDER									
MALE	3228	50.1(0.8) < 2%>	72.0(0.7)	13.4(0.6)	10.5(0.4)	2.9(0.4)	1.2(0.4)	0.1(0.1)	0.0
			274.4(1.8)	231.9(2.9)	243.1(3.0)	272.5(4.7)	255.2(10.3)	273.9(12.8)	0.0
FEMALE	3303	49.9(0.8) < 2%>	69.6(0.7)	16.3(0.6)	9.8(0.4)	2.6(0.5)	1.6(0.6)	0.1(0.1)	0.0
			271.3(1.4)	230.3(2.1)	239.1(2.5)	268.3(4.5)	249.5(7.8)	234.8(14.5)	0.0
RACE/ETHNICITY									
WHITE	4223	70.8(0.4) < 1%>	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
			272.9(1.4)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
BLACK	917	14.8(0.4) < 3%>	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
			0.0(0.0)	231.0(2.2)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
HISPANIC	1000	10.1(0.3) < 3%>	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
			0.0(0.0)	0.0(0.0)	241.2(2.1)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
ASIAN/PACIFIC AMERICAN	285	2.7(0.4) <14%>	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
			0.0(0.0)	0.0(0.0)	0.0(0.0)	270.5(4.0)	0.0(0.0)	0.0(0.0)	0.0
AMER IND/ALASKAN NATV	95	1.4(0.5) <36%>!	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0
			0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	251.9(8.5)	0.0(0.0)	0.0
PARENTAL EDUCATION									
LESS THAN H.S.	584	8.8(0.6) < 7%>	60.7(2.7)	15.4(1.8)	21.2(1.9)	1.3(0.5)	1.3(0.6)	0.1(0.1)	0.0
			247.9(2.5)	219.9(5.3)	236.6(3.4)	242.3(12.6)	246.3(23.3)	273.5(****)	0.0
GRADUATED H.S.	1540	24.8(0.8) < 3%>	71.4(1.3)	15.7(1.0)	9.8(0.6)	1.4(0.3)	1.6(0.6)	0.1(0.1)	0.0
			263.2(1.5)	223.6(2.7)	238.1(2.7)	261.0(9.2)	239.8(13.7)	218.8(10.2)	0.0
SOME EDUC AFTER H.S.	1195	18.7(0.8) < 4%>	72.2(1.5)	14.5(1.2)	9.6(0.9)	2.0(0.5)	1.7(0.6)	0.0(0.0)	0.0
			276.3(1.7)	240.8(3.7)	248.6(2.8)	275.1(5.8)	269.9(16.5)	0.0(0.0)	0.0
GRADUATED COLLEGE	2642	39.6(1.6) < 4%>	76.5(1.0)	13.1(0.8)	5.6(0.5)	3.7(0.5)	1.0(0.4)	0.2(0.1)	0.0
			284.2(1.6)	240.6(3.0)	254.0(3.5)	280.9(4.5)	259.8(4.4)	273.0(12.9)	0.0
UNKNOWN	559	7.9(0.4) < 5%>	48.8(1.9)	21.0(2.0)	22.8(1.9)	5.3(1.4)	1.9(0.7)	0.1(0.2)	0.0
			250.5(3.3)	213.4(4.6)	226.4(3.9)	245.3(6.7)	230.0(16.5)	226.2(****)	0.0

! INTERPRET WITH CAUTION: SAMPLING ERROR CANNOT BE ESTIMATED ACCURATELY SINCE COEFFICIENT OF VARIATION OF ESTIMATED NUMBER OF STUDENTS EXCEEDS 20%

Table 16-103

Weighted Response Percentages and Proficiency Means (with Standard Errors)
for Main Focused-BIB Science Samples, by Parents' Education Level, Grade 8

PARENTS' EDUCATION	N	WEIGHTED PCT <CV>	NOT HS	GRAD HS	POST HS	GRAD COL	UNKNOWN	MISSING
-- TOTAL --	6520	100.0(0.0) < 1%>	8.9(0.6) 241.1(2.3)	24.8(0.8) 254.1(1.3)	18.8(0.8) 268.4(1.4)	39.7(1.6) 276.4(1.7)	7.9(0.4) 236.5(2.3)	0.1 240.9
GENDER								
MALE	3221	50.0(0.8) < 2%>	7.1(0.7) 244.8(3.3)	24.9(1.0) 255.9(1.8)	17.3(0.8) 270.3(2.3)	41.7(1.7) 277.4(2.0)	9.0(0.6) 240.6(3.0)	0.2 250.2
FEMALE	3299	50.0(0.8) < 2%>	10.6(0.9) 238.7(2.8)	24.7(0.9) 252.3(1.5)	20.3(1.0) 266.7(1.5)	37.7(1.7) 275.3(1.9)	6.8(0.5) 231.0(2.7)	0.1 218.1
RACE/ETHNICITY								
WHITE	4220	70.9(0.4) < 1%>	7.6(0.7) 247.9(2.5)	25.0(1.0) 263.2(1.5)	19.1(1.0) 276.3(1.7)	42.9(1.9) 284.2(1.6)	5.4(0.3) 250.5(3.3)	0.1 267.1
BLACK	913	14.8(0.3) < 3%>	9.2(1.3) 219.9(5.3)	26.3(1.7) 223.6(2.7)	18.4(1.6) 240.8(3.7)	35.0(1.9) 240.6(3.0)	11.2(1.2) 213.4(4.6)	0.3 183.0
HISPANIC	996	10.1(0.3) < 3%>	18.6(1.7) 236.6(3.4)	24.0(1.2) 238.1(2.7)	17.8(1.7) 248.6(2.8)	21.9(1.9) 254.0(3.5)	17.8(1.5) 226.4(3.9)	0.5 268.1
ASIAN/PACIFIC AMERICAN	285	2.7(0.4) <14%>	4.1(1.7) 242.3(12.6)	12.4(2.0) 261.0(9.2)	14.0(2.7) 275.1(5.8)	54.0(4.3) 280.9(4.5)	15.5(2.8) 245.3(6.7)	0.0 0.0
AMER IND/ALASKAN NATV	95	1.4(0.5) <36%>!	8.2(2.0) 246.3(23.3)	29.4(3.6) 239.8(13.7)	23.1(2.9) 269.9(16.5)	28.4(5.0) 259.8(4.4)	10.9(5.4) 230.0(16.5)	0.0 0.0
PARENTAL EDUCATION								
LESS THAN H.S.	584	8.9(0.6) < 7%>	100.0(0.0) 241.1(2.3)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
GRADUATED H.S.	1540	24.8(0.8) < 3%>	0.0(0.0) 0.0(0.0)	100.0(0.0) 254.1(1.3)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
SOME EDUC AFTER H.S.	1195	18.8(0.8) < 4%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 268.4(1.4)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
GRADUATED COLLEGE	2642	39.7(1.6) < 4%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 276.4(1.7)	0.0(0.0) 0.0(0.0)	0.0 0.0
UNKNOWN	559	7.9(0.4) < 5%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 236.5(2.3)	0.0 0.0

498 ! INTERPRET WITH CAUTION: SAMPLING ERROR CANNOT BE ESTIMATED ACCURATELY SINCE COEFFICIENT OF VARIATION OF ESTIMATED NUMBER OF STUDENTS EXCEEDS 20% 499

Table 16-104

Weighted Response Percentages and Proficiency Means (with Standard Errors)
for Main Focused-BIB Science Samples, by Gender, Grade 12

GENDER OF SUBJECT

	N	WEIGHTED PCT <CV>	MALE	FEMALE	MISSING
-- TOTAL --	6337	100.0(0.0) < 1%>	48.2(0.8) 298.9(1.5)	51.8(0.8) 288.6(1.2)	0.0 0.0
GENDER					
MALE	3058	48.2(0.8) < 2%>	100.0(0.0) 298.9(1.5)	0.0(0.0) 0.0(0.0)	0.0 0.0
FEMALE	3279	51.8(0.8) < 2%>	0.0(0.0) 0.0(0.0)	100.0(0.0) 288.6(1.2)	0.0 0.0
RACE/ETHNICITY					
WHITE	4443	73.2(0.4) < 1%>	49.0(1.0) 307.1(1.5)	51.0(1.0) 298.0(1.3)	0.0 0.0
BLACK	872	14.2(0.5) < 3%>	43.3(2.2) 261.2(2.7)	56.7(2.2) 252.6(2.9)	0.0 0.0
HISPANIC	700	8.2(0.3) < 4%>	49.1(2.9) 277.7(3.1)	50.9(2.9) 267.5(3.5)	0.0 0.0
ASIAN/PACIFIC AMERICAN	263	3.6(0.2) < 5%>	46.0(2.3) 315.2(10.6)	54.0(2.3) 302.3(5.0)	0.0 0.0
AMER IND/ALASKAN NATV	51	0.7(0.2) <36%>!	56.9(7.7) 288.7(7.1)	43.1(7.7) 281.8(5.9)	0.0 0.0
PARENTAL EDUCATION					
LESS THAN H.S.	490	7.5(0.6) < 8%>	37.3(2.4) 276.5(4.3)	62.7(2.4) 264.6(2.9)	0.0 0.0
GRADUATED H.S.	1443	23.7(0.8) < 3%>	47.7(1.6) 283.3(2.0)	52.3(1.6) 274.8(1.6)	0.0 0.0
SOME EDUC AFTER H.S.	1602	26.0(0.8) < 3%>	48.7(1.6) 298.9(1.7)	51.3(1.6) 291.7(1.7)	0.0 0.0
GRADUATED COLLEGE	2656	40.4(1.3) < 3%>	49.9(1.0) 313.1(1.9)	50.1(1.0) 303.3(1.3)	0.0 0.0
UNKNOWN	130	2.1(0.2) <11%>	53.3(5.2) 260.5(6.8)	46.7(5.2) 235.9(6.6)	0.0 0.0

! INTERPRET WITH CAUTION: SAMPLING ERROR CANNOT BE ESTIMATED ACCURATELY SINCE COEFFICIENT OF VARIATION OF ESTIMATED NUMBER OF STUDENTS EXCEEDS 20%

Table 16-105

Weighted Response Percentages and Proficiency Means (with Standard Errors)
for Main Focused-BIB Science Samples, by Derived Race/Ethnicity, Grade 12

DERIVED RACE/ETHNICITY	N	WEIGHTED PCT <CV>	WHITE	BLACK	HISPANIC	ASIAN	AMER IND	UNCLASSIFIED	MISSING
-- TOTAL --	6337	100.0(0.0) < 1%>	73.2(0.4) 302.5(1.3)	14.2(0.5) 256.3(2.4)	8.2(0.3) 272.5(2.8)	3.6(0.2) 308.2(7.1)	0.7(0.2) 285.7(4.6)	0.1(0.0) 277.2(18.1)	0.0 0.0
GENDER									
MALE	3058	48.2(0.8) < 2%>	74.5(0.9) 307.1(1.5)	12.8(0.8) 261.2(2.7)	8.4(0.6) 277.7(3.1)	3.4(0.2) 315.2(10.6)	0.8(0.3) 288.7(7.1)	0.1(0.1) 271.7(35.7)	0.0 0.0
FEMALE	3279	51.8(0.8) < 2%>	72.0(0.8) 293.0(1.3)	15.6(0.7) 252.6(2.9)	8.1(0.5) 267.5(3.5)	3.7(0.3) 302.3(5.0)	0.6(0.3) 281.8(5.9)	0.1(0.1) 284.4(9.6)	0.0 0.0
RACE/ETHNICITY									
WHITE	4443	73.2(0.4) < 1%>	100.0(0.0) 302.5(1.3)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
BLACK	872	14.2(0.5) < 3%>	0.0(0.0) 0.0(0.0)	100.0(0.0) 256.3(2.4)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
HISPANIC	700	8.2(0.3) < 4%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 272.5(2.8)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
ASIAN/PACIFIC AMERICAN	263	3.6(0.2) < 5%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 308.2(7.1)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
AMER IND/ALASKAN NATV	51	0.7(0.2) <36%>!	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 285.7(4.6)	0.0(0.0) 0.0(0.0)	0.0 0.0
PARENTAL EDUCATION									
LESS THAN H.S.	490	7.5(0.6) < 8%>	52.9(3.0) 276.3(4.0)	16.6(2.2) 247.8(4.9)	27.2(2.8) 265.2(4.1)	2.6(0.8) 301.4(6.9)	0.5(0.4) 271.5(11.5)	0.2(0.2) 209.5(****)	0.0 0.0
GRADUATED H.S.	1443	23.7(0.8) < 3%>	71.7(1.2) 288.7(1.5)	17.4(1.0) 243.9(3.5)	7.9(0.5) 266.0(3.4)	2.1(0.5) 290.5(6.2)	0.8(0.4) 262.9(9.2)	0.0(0.0) 264.7(9.3)	0.0 0.0
SOME EDUC AFTER H.S.	1602	26.0(0.8) < 3%>	74.9(1.5) 302.3(1.5)	14.7(1.0) 267.0(3.4)	6.9(0.6) 278.9(3.4)	2.5(1.0) 294.9(12.1)	1.0(0.4) 295.7(8.2)	0.0(0.0) 0.0(0.0)	0.0 0.0
GRADUATED COLLEGE	2656	40.4(1.3) < 3%>	78.7(1.0) 314.5(1.2)	10.7(0.8) 266.1(3.1)	4.7(0.5) 287.1(3.9)	5.2(0.4) 319.7(6.4)	0.5(0.2) 298.1(10.4)	0.2(0.1) 294.6(8.4)	0.0 0.0
UNKNOWN	130	2.1(0.2) <11%>	39.3(6.1) 258.3(7.9)	29.6(5.5) 226.8(9.8)	26.0(4.7) 253.9(12.0)	3.9(1.5) 286.2(11.5)	1.1(1.2) 267.5(2.1)	0.2(0.2) 261.1(23.1)	0.0 0.0

! INTERPRET WITH CAUTION: SAMPLING ERROR CANNOT BE ESTIMATED ACCURATELY SINCE COEFFICIENT OF VARIATION OF ESTIMATED NUMBER OF STUDENTS EXCEEDS 20%

Table 16-106

Weighted Response Percentages and Proficiency Means (with Standard Errors)
for Main Focused-BIB Science Samples, by Parents' Education Level, Grade 12

PARENT'S EDUCATION LEVEL	N	WEIGHTED PCT <CV>	NOT HS	GRAD HS	POST HS	GRAD COL	UNKNOWN	MISSING
-- TOTAL --	6321	100.0(0.0) < 1%>	7.5(0.6) 269.0(2.5)	23.8(0.8) 278.9(1.3)	26.1(0.8) 295.2(1.3)	40.5(1.3) 308.2(1.4)	2.1(0.2) 249.0(5.3)	0.2 234.4
GENDER								
MALE	3050	48.2(0.8) < 2%>	5.8(0.6) 276.5(4.3)	23.5(1.0) 283.3(2.0)	26.4(1.1) 298.9(1.7)	41.9(1.6) 313.1(1.9)	2.3(0.4) 260.5(6.8)	0.3 235.9
FEMALE	3271	51.8(0.8) < 2%>	9.1(0.8) 264.6(2.9)	24.0(0.9) 274.8(1.6)	25.8(0.9) 291.7(1.7)	39.2(1.4) 303.3(1.3)	1.9(0.3) 235.9(6.6)	0.2 232.9
RACE/ETHNICITY								
WHITE	4439	73.3(0.4) < 1%>	5.4(0.6) 276.3(4.0)	23.3(0.9) 288.7(1.5)	26.7(1.1) 302.3(1.5)	43.5(1.7) 314.5(1.2)	1.1(0.2) 258.3(7.9)	0.1 260.3
BLACK	867	14.2(0.5) < 3%>	8.8(1.3) 247.8(4.9)	29.3(1.5) 243.9(3.5)	27.1(1.6) 267.0(3.4)	30.5(1.8) 266.1(3.1)	4.4(0.9) 226.8(9.8)	0.6 225.1
HISPANIC	694	8.2(0.3) < 4%>	25.0(2.7) 265.2(4.1)	23.1(1.7) 266.0(3.4)	22.1(1.7) 278.9(3.4)	23.2(2.5) 287.1(3.9)	6.6(1.3) 253.9(12.0)	0.6 230.0
ASIAN/PACIFIC AMERICAN	262	3.6(0.2) < 5%>	5.5(1.7) 301.4(6.9)	14.4(3.4) 290.5(6.2)	18.6(7.0) 294.9(12.1)	59.2(4.7) 319.7(6.4)	2.3(0.9) 286.2(11.5)	0.9 207.6
AMER IND/ALASKAN NATV	51	0.7(0.2) <36%>!	5.0(2.9) 271.5(11.5)	26.0(5.6) 262.9(9.2)	35.6(6.7) 295.7(8.2)	30.0(6.7) 298.1(10.4)	3.4(3.4) 267.5(2.1)	0.0 0.0
PARENTAL EDUCATION								
LESS THAN H.S.	490	7.5(0.6) < 8%>	100.0(0.0) 269.0(2.5)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
GRADUATED H.S.	1443	23.8(0.8) < 3%>	0.0(0.0) 0.0(0.0)	100.0(0.0) 278.9(1.3)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
SOME EDUC AFTER H.S.	1602	26.1(0.8) < 3%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 295.2(1.3)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0 0.0
GRADUATED COLLEGE	2656	40.5(1.3) < 3%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 308.2(1.4)	0.0(0.0) 0.0(0.0)	0.0 0.0
UNKNOWN	130	2.1(0.2) <11%>	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	0.0(0.0) 0.0(0.0)	100.0(0.0) 249.0(5.3)	0.0 0.0

! INTERPRET WITH CAUTION: SAMPLING ERROR CANNOT BE ESTIMATED ACCURATELY SINCE COEFFICIENT OF VARIATION OF ESTIMATED NUMBER OF STUDENTS EXCEEDS 20%

Table 16-107

Weighted Response Percentages and Writing Meanparts Scores (with Standard Errors) for Bridge to 1984, by Gender, Grade 4

GENDER OF SUBJECT		N	WTD % (CV)	MALE	FEMALE	MISSING
--TOTAL--	%	4367	100.0 (1%)	50.8(0.9)	49.2(0.9)	0.0(0.0)
	M	4367	100.0 (1%)	173 6(1.6)	192.5(2.2)	0.0(0.0)
GENDER						
MALE	%	2188	50.8 (2%)	100.0(0.0)	0.0(0.0)	0.0(0.0)
	M	2188	50.8 (2%)	173.6(1.6)	0.0(0.0)	0.0(0.0)
FEMALE	%	2179	49.2 (2%)	0.0(0.0)	100.0(0.0)	0.0(0.0)
	M	2179	49.2 (2%)	0.0(0.0)	192.5(2.2)	0.0(0.0)
RACE/ETHNICITY						
WHITE	%	2757	70.4 (1%)	50.3(1.1)	49.7(1.1)	0.0(0.0)
	M	2757	70.4 (1%)	181.5(1.5)	199.7(2.5)	0.0(0.0)
BLACK	%	686	15.2 (2%)	52.2(1.7)	47.8(1.7)	0.0(0.0)
	M	686	15.2 (2%)	143.2(5.2)	166.4(5.4)	0.0(0.0)
HISPANIC	%	696	10.7 (2%)	52.0(2.2)	48.0(2.2)	0.0(0.0)
	M	696	10.7 (2%)	160.3(3.4)	175.8(4.9)	0.0(0.0)
OTHER	%	228	3.8 (2%)	51.5(4.6)	48.5(4.6)	0.0(0.0)
	M	228	3.8 (2%)	178.1(6.4)	200.8(8.3)	0.0(0.0)
PARENTAL EDUCATION						
LESS THAN HS	%	228	5.3 (8%)	42.7(3.8)	57.3(3.8)	0.0(0.0)
	M	228	5.3 (8%)	155.5(5.9)	176.3(6.4)	0.0(0.0)
GRADUATED HS	%	791	18.6 (4%)	50.2(1.8)	49.8(1.8)	0.0(0.0)
	M	791	18.6 (4%)	172.4(3.5)	193.2(3.6)	0.0(0.0)
SOME EDU AFTER HS	%	214	4.7 (8%)	50.7(3.6)	49.3(3.6)	0.0(0.0)
	M	214	4.7 (8%)	178.7(6.2)	210.5(9.1)	0.0(0.0)
GRADUATED COLLEGE	%	1721	39.3 (4%)	54.0(1.5)	46.0(1.5)	0.0(0.0)
	M	1721	39.3 (4%)	181.2(3.2)	201.8(2.8)	0.0(0.0)
UNKNOWN	%	1399	31.7 (4%)	48.7(1.7)	51.3(1.7)	0.0(0.0)
	M	1399	31.7 (4%)	166.7(2.7)	181.2(2.8)	0.0(0.0)

% = PERCENTAGE; M = MEANPARTS TIMES 100

507

506

Table 16-108

Weighted Response Percentages and Writing Meanparts Scores (with Standard Errors)
for Bridge to 1984, by Derived Race/Ethnicity, Grade 4

DERIVED RACE/ETHNICITY		N	WTD % (CV)	WHITE	BLACK	HISPANIC	ASIAN	AMER IND	UNCLASSI	MISSING
--TOTAL--	%	4367	100.0 (1%)	70.4(0.3)	15.2(0.3)	10.7(0.2)	2.3(0.4)	1.4(0.4)	0.1(0.1)	0.0(0.0)
	M	4367	100.0 (1%)	190.9(1.6)	155.0(4.8)	167.8(3.4)	188.9(9.5)	187.4(7.7)	200.7(13.4)	0.0(0.0)
GENDER										
MALE	%	2188	50.8 (2%)	69.7(0.7)	15.6(0.5)	10.9(0.4)	2.2(0.3)	1.6(0.5)	0.0(0.0)	0.0(0.0)
	M	2188	50.8 (2%)	181.5(1.5)	143.2(5.2)	160.3(3.4)	176.7(11.2)	178.8(11.0)	175.0(0.0)	0.0(0.0)
FEMALE	%	2179	49.2 (2%)	71.1(0.8)	14.8(0.6)	10.4(0.5)	2.3(0.6)	1.3(0.4)	0.1(0.1)	0.0(0.0)
	M	2179	49.2 (2%)	199.7(2.5)	166.4(5.4)	175.8(4.9)	198.8(12.5)	202.7(13.5)	*****(****)	0.0(0.0)
RACE/ETHNICITY										
WHITE	%	2757	70.4 (1%)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
	M	2757	70.4 (1%)	190.9(1.6)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
BLACK	%	686	15.2 (2%)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
	M	686	15.2 (2%)	0.0(0.0)	155.0(4.8)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
HISPANIC	%	696	10.7 (2%)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
	M	696	10.7 (2%)	0.0(0.0)	0.0(0.0)	167.8(3.4)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
OTHER	%	228	3.8 (2%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	60.1(10.9)	37.9(10.5)	2.0(1.4)	0.0(0.0)
	M	228	3.8 (2%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	188.9(9.5)	187.4(7.7)	200.7(13.4)	0.0(0.0)
PARENTAL EDUCATION										
LESS THAN HS	%	228	5.3 (8%)	67.9(3.0)	13.8(3.0)	14.6(1.9)	1.1(1.1)	2.7(1.0)	0.0(0.0)	0.0(0.0)
	M	228	5.3 (8%)	171.9(5.5)	164.0(25.8)	154.0(10.3)	*****(****)	168.5(25.2)	0.0(0.0)	0.0(0.0)
GRADUATED HS	%	791	18.6 (4%)	72.4(1.4)	16.8(1.2)	9.2(1.2)	0.6(0.4)	1.0(0.4)	0.0(0.0)	0.0(0.0)
	M	791	18.6 (4%)	191.1(3.3)	146.5(5.7)	181.9(7.3)	164.2(97.5)	181.9(58.7)	0.0(0.0)	0.0(0.0)
SOME EDU AFTER HS	%	214	4.7 (8%)	70.7(2.9)	15.2(2.3)	8.9(1.6)	3.1(1.2)	2.1(1.0)	0.0(0.0)	0.0(0.0)
	M	214	4.7 (8%)	200.7(6.9)	156.1(13.3)	190.4(18.6)	191.7(23.0)	261.3(5.1)	0.0(0.0)	0.0(0.0)
GRADUATED COLLEGE	%	1721	39.3 (4%)	72.0(1.2)	15.0(1.1)	9.0(0.5)	2.8(0.4)	1.1(0.4)	0.1(0.1)	0.0(0.0)
	M	1721	39.3 (4%)	200.0(2.0)	162.4(6.6)	165.4(4.2)	196.4(12.0)	183.2(13.6)	175.0(0.0)	0.0(0.0)
UNKNOWN	%	1399	31.7 (4%)	67.8(1.3)	14.5(1.2)	13.1(0.8)	2.7(0.4)	1.8(0.5)	0.1(0.1)	0.0(0.0)
	M	1399	31.7 (4%)	180.4(2.3)	150.4(7.3)	164.1(4.7)	185.4(6.1)	186.4(15.0)	*****(****)	0.0(0.0)

% = PERCENTAGE; M = MEANPARTS TIMES 100
*****(****) THE SAMPLE SIZE IS INSUFFICIENT TO PROVIDE RELIABLE ESTIMATES

509

508

Table 16-109

Weighted Response Percentages and Writing Meanparts Scores (with Standard Errors)
for Bridge to 1984, by Parents' Education Level, Grade 4

PARENTS' EDUCATION LEVEL		N	WTD % (CV)	NOT HS	GRAD HS	POST HS	GRAD COL	UNKNOWN	MISSING
--TOTAL--	%	4353	100.0 (1%)	5.3(0.4)	18.7(0.8)	4.7(0.4)	39.4(1.5)	31.9(1.1)	0.0(0.0)
	M	4353	100.0 (1%)	169.1(4.9)	183.0(2.8)	194.5(5.9)	191.3(2.3)	174.4(2.2)	0.0(0.0)
GENDER									
MALE	%	2181	50.8 (2%)	4.5(0.6)	18.4(1.0)	4.7(0.5)	41.8(1.7)	30.5(1.5)	0.0(0.0)
	M	2181	50.8 (2%)	155.5(5.9)	172.4(3.5)	178.7(6.2)	181.2(3.2)	166.7(2.7)	0.0(0.0)
FEMALE	%	2172	49.2 (2%)	6.2(0.7)	18.9(1.0)	4.8(0.5)	36.9(1.6)	33.2(1.4)	0.0(0.0)
	M	2172	49.2 (2%)	176.3(6.4)	193.2(3.6)	210.5(9.1)	201.8(2.8)	181.2(2.8)	0.0(0.0)
RACE/ETHNICITY									
WHITE	%	2752	70.5 (1%)	5.1(0.5)	19.2(0.9)	4.8(0.5)	40.2(1.6)	30.7(1.2)	0.0(0.0)
	M	2752	70.5 (1%)	171.9(5.5)	191.1(3.3)	200.7(6.9)	200.0(2.0)	180.4(2.3)	0.0(0.0)
BLACK	%	682	15.1 (2%)	4.9(1.2)	20.7(1.8)	4.8(0.8)	39.1(3.3)	30.6(3.1)	0.0(0.0)
	M	682	15.1 (2%)	164.0(25.8)	146.5(5.7)	156.1(13.3)	162.4(6.6)	150.4(7.3)	0.0(0.0)
HISPANIC	%	692	10.6 (2%)	7.3(1.0)	16.2(2.1)	4.0(0.7)	33.4(1.8)	39.1(2.5)	0.0(0.0)
	M	692	10.6 (2%)	154.0(10.3)	181.9(7.3)	190.4(18.6)	165.4(4.2)	164.1(4.7)	0.0(0.0)
OTHER	%	227	3.7 (2%)	5.3(1.2)	7.9(1.8)	6.5(1.6)	41.4(3.3)	38.8(3.3)	0.0(0.0)
	M	227	3.7 (2%)	172.9(34.3)	171.8(21.0)	206.8(26.8)	193.0(6.9)	185.2(6.2)	0.0(0.0)
PARENTAL EDUCATION									
LESS THAN HS	%	228	5.3 (8%)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
	M	228	5.3 (8%)	169.1(4.9)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
GRADUATED HS	%	791	18.7 (4%)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
	M	791	18.7 (4%)	0.0(0.0)	183.0(2.8)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
SOME EDU AFTER HS	%	214	4.7 (8%)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
	M	214	4.7 (8%)	0.0(0.0)	0.0(0.0)	194.5(5.9)	0.0(0.0)	0.0(0.0)	0.0(0.0)
GRADUATED COLLEGE	%	1721	39.4 (4%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)
	M	1721	39.4 (4%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	191.3(2.3)	0.0(0.0)	0.0(0.0)
UNKNOWN	%	1399	31.8 (4%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)
	M	1399	31.8 (4%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	174.4(2.2)	0.0(0.0)

% = PERCENTAGE; M = MEANPARTS TIMES 100

Table 16-110

Weighted Response Percentages and Writing Meanparts Scores (with Standard Errors)
for Bridge to 1984, by Gender, Grade 8

GENDER OF SUBJECT		N	WTD % (CV)	MALE	FEMALE	MISSING
--TOTAL--	%	4665	100.0 (1%)	50.5(0.9)	49.5(0.9)	0.0(0.0)
	M	4665	100.0 (1%)	186.8(1.6)	208.3(1.5)	0.0(0.0)
GENDER						
MALE	%	2346	50.5 (2%)	100.0(0.0)	0.0(0.0)	0.0(0.0)
	M	2346	50.5 (2%)	186.8(1.6)	0.0(0.0)	0.0(0.0)
FEMALE	%	2319	49.5 (2%)	0.0(0.0)	100.0(0.0)	0.0(0.0)
	M	2319	49.5 (2%)	0.0(0.0)	208.3(1.5)	0.0(0.0)
RACE/ETHNICITY						
WHITE	%	3237	70.2 (0%)	49.8(1.0)	50.2(1.0)	0.0(0.0)
	M	3237	70.2 (0%)	190.6(2.0)	213.3(1.7)	0.0(0.0)
BLACK	%	566	15.1 (1%)	49.0(2.4)	51.0(2.4)	0.0(0.0)
	M	566	15.1 (1%)	173.4(4.1)	189.2(3.4)	0.0(0.0)
HISPANIC	%	616	10.3 (2%)	54.9(2.5)	45.1(2.5)	0.0(0.0)
	M	616	10.3 (2%)	178.9(3.7)	200.4(3.0)	0.0(0.0)
OTHER	%	246	4.4 (2%)	55.8(3.5)	44.2(3.5)	0.0(0.0)
	M	246	4.4 (2%)	186.7(5.1)	205.4(5.1)	0.0(0.0)
PARENTAL EDUCATION						
LESS THAN HS	%	342	7.6 (8%)	40.1(2.5)	59.9(2.5)	0.0(0.0)
	M	342	7.6 (8%)	175.5(5.4)	201.6(3.8)	0.0(0.0)
GRADUATED HS	%	1449	32.7 (4%)	49.6(1.6)	50.4(1.6)	0.0(0.0)
	M	1449	32.7 (4%)	185.1(2.3)	204.4(2.4)	0.0(0.0)
SOME EDU AFTER HS	%	533	11.7 (6%)	44.4(2.5)	55.6(2.5)	0.0(0.0)
	M	533	11.7 (6%)	192.9(4.7)	216.7(2.8)	0.0(0.0)
GRADUATED COLLEGE	%	1887	38.1 (4%)	53.5(1.2)	46.5(1.2)	0.0(0.0)
	M	1887	38.1 (4%)	193.8(2.4)	214.6(2.6)	0.0(0.0)
UNKNOWN	%	437	9.6 (6%)	55.4(2.3)	44.6(2.3)	0.0(0.0)
	M	437	9.6 (6%)	166.3(4.1)	189.1(4.3)	0.0(0.0)

% = PERCENTAGE; M = MEANPARTS TIMES 100

512

513

Table 16-111

Weighted Response Percentages and Writing Meanparts Scores (with Standard Errors) for Bridge to 1984, by Derived Race/Ethnicity, Grade 8

DERIVED RACE/ETHNICITY		N	WTD % (CV)	WHITE	BLACK	HISPANIC	ASIAN	AMER IND	UNCLASSI	MISSING
--TOTAL--	%	4665	100.0 (1%)	70.2(0.2)	15.1(0.2)	10.3(0.2)	2.6(0.5)	1.7(0.5)	0.1(0.0)	0.0(0.0)
	M	4665	100.0 (1%)	202.3(1.5)	182.1(2.8)	189.0(3.0)	20' /(9.4)	185.2(6.2)	192.6(43.3)	0.0(0.0)
GENDER										
MALE	%	2346	50.5 (2%)	69.3(0.8)	14.7(0.7)	11.2(0.5)	2.7(0.4)	2.0(0.6)	0.1(0.1)	0.0(0.0)
	M	2346	50.5 (2%)	190.6(2.0)	173.4(4.1)	178.9(3.7)	197.9(9.8)	171.5(7.6)	*****(****)	0.0(0.0)
FEMALE	%	2319	49.5 (2%)	71.2(0.8)	15.5(0.6)	9.4(0.5)	2.5(0.6)	1.3(0.4)	0.1(0.0)	0.0(0.0)
	M	2319	49.5 (2%)	213.3(1.7)	189.2(3.4)	200.4(3.0)	205.6(8.8)	203.3(7.1)	*****(****)	0.0(0.0)
RACE/ETHNICITY										
WHITE	%	3237	70.2 (0%)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
	M	3237	70.2 (0%)	202.3(1.5)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
BLACK	%	566	15.1 (1%)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
	M	566	15.1 (1%)	0.0(0.0)	182.1(2.8)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
HISPANIC	%	616	10.3 (2%)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
	M	616	10.3 (2%)	0.0(0.0)	0.0(0.0)	189.0(3.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
OTHER	%	246	4.4 (2%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	60.3(10.4)	38.1(10.5)	1.6(0.9)	0.0(0.0)
	M	246	4.4 (2%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	201.9(9.4)	185.2(6.2)	192.6(43.3)	0.0(0.0)
PARENTAL EDUCATION										
LESS THAN HS	%	342	7.6 (8%)	60.4(3.4)	9.9(2.1)	25.8(2.9)	1.0(0.5)	2.9(1.0)	0.0(0.0)	0.0(0.0)
	M	342	7.6 (8%)	193.4(4.3)	168.4(13.6)	197.9(6.6)	180.0(0.0)	180.7(18.6)	0.0(0.0)	0.0(0.0)
GRADUATED HS	%	1449	32.7 (4%)	72.5(1.5)	16.2(1.4)	7.8(0.7)	1.8(0.9)	1.7(0.5)	0.1(0.1)	0.0(0.0)
	M	1449	32.7 (4%)	199.8(2.2)	181.6(4.0)	179.1(5.4)	182.4(11.3)	187.7(8.2)	100.0(0.0)	0.0(0.0)
SOME EDU AFTER HS	%	533	11.7 (6%)	72.8(2.3)	15.8(1.7)	8.5(1.4)	1.0(0.4)	1.9(0.7)	0.0(0.0)	0.0(0.0)
	M	533	11.7 (6%)	210.5(2.6)	187.1(10.1)	199.2(6.2)	226.1(10.5)	184.6(27.0)	0.0(0.0)	0.0(0.0)
GRADUATED COLLEGE	%	1887	38.1 (4%)	74.0(1.2)	15.0(1.1)	6.0(0.7)	3.5(0.3)	1.4(0.6)	0.1(0.1)	0.0(0.0)
	M	1887	38.1 (4%)	207.6(2.0)	184.4(4.2)	203.2(4.9)	212.3(6.3)	192.0(9.2)	*****(****)	0.0(0.0)
UNKNOWN	%	437	9.6 (6%)	52.7(2.4)	14.8(2.4)	25.7(2.4)	5.4(1.4)	1.4(0.8)	0.0(0.0)	0.0(0.0)
	M	437	9.6 (6%)	179.4(4.4)	173.1(5.4)	172.7(4.3)	189.8(10.0)	139.3(33.2)	0.0(0.0)	0.0(0.0)

% = PERCENTAGE; M = MEANPARTS TIMES 100
*****(****) THE SAMPLE SIZE IS INSUFFICIENT TO PROVIDE RELIABLE ESTIMATES

515

511

Table 16-112

Weighted Response Percentages and Writing Meanparts Scores (with Standard Errors)
for Bridge to 1984, by Parents' Education Level, Grade 8

PARENTS' EDUCATION LEVEL		N	WTD % (CV)	NOT HS	GRAD HS	POST HS	GRAD COL	UNKNOWN	MISSING
--TOTAL--	%	4648	100.0 (1%)	7.6(0.6)	32.8(1.1)	11.7(0.7)	38.3(1.5)	9.6(0.6)	0.0(0.0)
	M	4648	100.0 (1%)	191.7(3.7)	195.1(1.9)	206.9(2.7)	203.7(2.0)	177.1(3.0)	0.0(0.0)
GENDER									
MALE	%	2333	50.3 (2%)	6.1(0.6)	32.3(1.5)	10.3(0.9)	40.7(1.7)	10.6(0.8)	0.0(0.0)
	M	2333	50.3 (2%)	175.5(5.4)	185.1(2.3)	192.9(4.7)	193.8(2.4)	166.3(4.1)	0.0(0.0)
FEMALE	%	2315	49.7 (2%)	9.2(0.8)	33.3(1.3)	13.1(0.8)	35.8(1.7)	8.6(0.7)	0.0(0.0)
	M	2315	49.7 (2%)	201.6(3.8)	204.4(2.4)	216.7(2.8)	214.6(2.6)	189.1(4.3)	0.0(0.0)
RACE/ETHNICITY									
WHITE	%	3228	70.3 (1%)	6.6(0.6)	33.9(1.2)	12.1(0.8)	40.3(1.8)	7.2(0.5)	0.0(0.0)
	M	3228	70.3 (1%)	193.4(4.3)	199.8(2.2)	210.5(2.6)	207.6(2.0)	179.4(4.4)	0.0(0.0)
BLACK	%	564	15.1 (2%)	5.0(1.1)	35.2(3.5)	12.2(1.7)	38.2(3.0)	9.4(1.8)	0.0(0.0)
	M	564	15.1 (2%)	168.4(13.6)	181.6(4.0)	187.1(10.1)	184.4(4.2)	173.1(5.4)	0.0(0.0)
HISPANIC	%	611	10.3 (2%)	19.1(2.8)	24.8(2.2)	9.7(1.6)	22.5(2.8)	23.9(2.5)	0.0(0.0)
	M	611	10.3 (2%)	197.9(6.6)	179.1(5.4)	199.2(6.2)	203.2(4.9)	172.7(4.3)	0.0(0.0)
OTHER	%	245	4.4 (2%)	6.8(1.5)	26.9(4.9)	7.8(1.9)	43.5(4.8)	14.9(3.1)	0.0(0.0)
	M	245	4.4 (2%)	178.6(12.2)	184.0(6.8)	196.1(15.7)	206.4(4.6)	178.9(8.2)	0.0(0.0)
PARENTAL EDUCATION									
LESS THAN HS	%	342	7.6 (8%)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
	M	342	7.6 (8%)	191.7(3.7)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
GRADUATED HS	%	1449	32.8 (4%)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
	M	1449	32.8 (4%)	0.0(0.0)	195.1(1.9)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
SOME EDU AFTER HS	%	533	11.7 (6%)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
	M	533	11.7 (6%)	0.0(0.0)	0.0(0.0)	206.9(2.7)	0.0(0.0)	0.0(0.0)	0.0(0.0)
GRADUATED COLLEGE	%	1887	38.3 (4%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)
	M	1887	38.3 (4%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	203.7(2.0)	0.0(0.0)	0.0(0.0)
UNKNOWN	%	437	9.6 (6%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)
	M	437	9.6 (6%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	177.1(3.0)	0.0(0.0)

% = PERCENTAGE; M = MEANPARTS TIMES 100

516

517

Table 16-113

Weighted Response Percentages and Writing Meanparts Scores (with Standard Errors) for Bridge to 1984, by Gender, Grade 12

GENDER OF SUBJECT		N	WTD % (CV)	MALE	FEMALE	MISSING
--TOTAL--	%	4385	100.0 (0%)	50.9(0.8)	49.1(0.8)	0.0(0.0)
	M	4385	100.0 (0%)	199.8(2.0)	223.7(1.4)	0.0(0.0)
GENDER						
MALE	%	2206	50.9 (2%)	100.0(0.0)	0.0(0.0)	0.0(0.0)
	M	2206	50.9 (2%)	199.8(2.0)	0.0(0.0)	0.0(0.0)
FEMALE	%	2179	49.1 (2%)	0.0(0.0)	100.0(0.0)	0.0(0.0)
	M	2179	49.1 (2%)	0.0(0.0)	223.7(1.4)	0.0(0.0)
RACE/ETHNICITY						
WHITE	%	3222	71.2 (0%)	51.2(0.9)	48.8(0.9)	0.0(0.0)
	M	3222	71.2 (0%)	204.2(2.0)	229.8(1.7)	0.0(0.0)
BLACK	%	601	15.5 (1%)	46.7(2.1)	53.3(2.1)	0.0(0.0)
	M	601	15.5 (1%)	181.7(4.1)	204.9(2.6)	0.0(0.0)
HISPANIC	%	342	8.7 (1%)	53.4(3.0)	46.6(3.0)	0.0(0.0)
	M	342	8.7 (1%)	191.1(5.2)	206.8(4.8)	0.0(0.0)
OTHER	%	220	4.5 (1%)	56.6(3.8)	43.4(3.8)	0.0(0.0)
	M	220	4.5 (1%)	202.0(6.0)	221.7(6.0)	0.0(0.0)
PARENTAL EDUCATION						
LESS THAN HS	%	345	8.5 (6%)	45.7(3.0)	54.3(3.0)	0.0(0.0)
	M	345	8.5 (6%)	167.4(6.4)	209.8(4.3)	0.0(0.0)
GRADUATED HS	%	1285	29.5 (4%)	51.7(1.2)	48.3(1.2)	0.0(0.0)
	M	1285	29.5 (4%)	193.6(3.3)	217.1(2.4)	0.0(0.0)
SOME EDU AFTER HS	%	826	18.4 (3%)	46.8(2.1)	53.2(2.1)	0.0(0.0)
	M	826	18.4 (3%)	207.5(2.9)	222.2(3.3)	0.0(0.0)
GRADUATED COLLEGE	%	1807	40.4 (4%)	52.4(1.3)	47.6(1.3)	0.0(0.0)
	M	1807	40.4 (4%)	209.4(2.3)	234.1(2.5)	0.0(0.0)
UNKNOWN	%	113	3.0 (9%)	62.4(5.4)	37.6(5.4)	0.0(0.0)
	M	113	3.0 (9%)	173.8(6.9)	191.0(11.7)	0.0(0.0)

% = PERCENTAGE; M = MEANPARTS TIMES 100

518

519

Table 16-114

Weighted Response Percentages and Writing Meanparts Scores (with Stan ... d Errors)
for Bridge to 1984, by Derived Race/Ethnicity, Grade 12

DERIVED RACE/ETHNICITY		N	WTD % (CV)	WHITE	BLACK	HISPANIC	ASIAN	AMER IND	UNCLASSI	MISSING
--TOTAL--	%	4385	100.0 (0%)	71.2(0.2)	15.5(0.2)	8.7(0.1)	3.0(0.4)	1.1(0.3)	0.4(0.2)	0.0(0.0)
	M	4385	100.0 (0%)	216.9(1.5)	193.7(2.3)	197.6(3.9)	209.1(5.9)	208.1(7.8)	232.9(9.0)	0.0(0.0)
GENDER										
MALE	%	2206	50.9 (2%)	71.6(0.8)	14.2(0.6)	9.2(0.5)	3.1(0.4)	1.4(0.4)	0.5(0.3)	0.0(0.0)
	M	2206	50.9 (2%)	204.2(2.0)	181.7(4.1)	191.1(5.2)	199.0(8.5)	204.7(7.9)	217.6(80.3)	0.0(0.0)
FEMALE	%	2179	49.1 (2%)	70.9(0.7)	16.8(0.7)	8.3(0.5)	3.0(0.4)	0.8(0.3)	0.2(0.1)	0.0(0.0)
	M	2179	49.1 (2%)	229.8(1.7)	204.9(2.6)	206.8(4.8)	219.7(7.3)	209.6(10.6)	279.3(6.4)	0.0(0.0)
RACE/ETHNICITY										
WHITE	%	3222	71.2 (0%)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
	M	3222	71.2 (0%)	216.9(1.5)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
BLACK	%	601	15.5 (1%)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
	M	601	15.5 (1%)	0.0(0.0)	193.7(2.3)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
HISPANIC	%	342	8.7 (1%)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
	M	342	8.7 (1%)	0.0(0.0)	0.0(0.0)	197.6(3.9)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
OTHER	%	220	4.5 (1%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	67.3(8.2)	24.5(6.7)	8.2(4.3)	0.0(0.0)
	M	220	4.5 (1%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	209.1(5.9)	208.1(7.8)	232.9(9.0)	0.0(0.0)
PARENTAL EDUCATION										
LESS THAN HS	%	345	8.5 (6%)	55.6(2.8)	15.3(2.0)	26.0(2.4)	1.9(0.8)	1.2(0.6)	0.0(0.0)	0.0(0.0)
	M	345	8.5 (6%)	196.2(4.3)	179.5(10.8)	181.8(4.4)	171.8(15.8)	145.1(30.0)	0.0(0.0)	0.0(0.0)
GRADUATED HS	%	1285	29.5 (4%)	71.3(1.3)	17.0(1.1)	7.5(0.9)	2.8(1.0)	1.0(0.4)	0.4(0.2)	0.0(0.0)
	M	1285	29.5 (4%)	210.2(2.9)	188.4(4.0)	199.6(7.2)	184.9(10.2)	218.4(8.8)	203.1(24.7)	0.0(0.0)
SOME EDU AFTER HS	%	826	18.4 (3%)	72.7(1.3)	15.7(0.9)	7.9(1.0)	2.4(0.4)	1.1(0.4)	0.3(0.2)	0.0(0.0)
	M	826	18.4 (3%)	219.3(2.3)	199.5(5.3)	207.2(7.9)	208.6(10.6)	199.4(38.1)	182.7(18.4)	0.0(0.0)
GRADUATED COLLEGE	%	1807	40.4 (4%)	76.7(1.1)	13.0(0.9)	5.0(0.6)	3.7(0.4)	1.2(0.4)	0.4(0.2)	0.0(0.0)
	M	1807	40.4 (4%)	224.2(1.9)	200.7(5.8)	219.3(7.8)	226.6(7.3)	223.2(14.3)	224.8(36.8)	0.0(0.0)
UNKNOWN	%	113	3.0 (9%)	32.8(4.8)	34.1(6.1)	27.0(4.4)	4.4(1.6)	0.5(0.5)	1.3(1.3)	0.0(0.0)
	M	113	3.0 (9%)	187.7(12.8)	184.5(9.4)	165.7(12.4)	225.6(76.1)	100.0(0.0)	*****(****)	0.0(0.0)

% = PERCENTAGE; M = MEANPARTS TIMES 100
*****(****) THE SAMPLE SIZE IS INSUFFICIENT TO PROVIDE RELIABLE ESTIMATES

520

BEST COPY AVAILABLE

521

Table 16-115

Weighted Response Percentages and Writing Meanparts Scores (with Standard Errors) for Main Focused-BIB Writing Samples, by Parents' Education Level, Grade 12

PARENTS' EDUCATION LEVEL		N	WTD % (CV)	NOT HS	GRAD HS	POST HS	GRAD COL	UNKNOWN	MISSING
--TOTAL--	%	4376	100.0 (0%)	8.5(0.5)	29.6(1.1)	18.5(0.6)	40.5(1.4)	3.0(0.3)	0.0(0.0)
	M	4376	100.0 (0%)	190.2(3.3)	204.8(2.3)	215.1(2.3)	221.3(1.8)	180.7(5.8)	0.0(0.0)
GENDER									
MALE	%	2199	50.9 (2%)	7.6(0.7)	30.1(1.3)	17.0(0.9)	41.7(1.7)	3.6(0.4)	0.0(0.0)
	M	2199	50.9 (2%)	167.4(6.4)	193.6(3.3)	207.5(2.9)	209.4(2.3)	173.8(6.9)	0.0(0.0)
FEMALE	%	2177	49.1 (2%)	9.4(0.7)	29.1(1.3)	20.0(1.0)	39.3(1.6)	2.3(0.4)	0.0(0.0)
	M	2177	49.1 (2%)	209.8(4.3)	217.1(2.4)	222.2(3.3)	234.1(2.5)	191.0(11.7)	0.0(0.0)
RACE/ETHNICITY									
WHITE	%	3216	71.3 (0%)	6.6(0.5)	29.6(1.2)	18.8(0.6)	43.6(1.7)	1.4(0.2)	0.0(0.0)
	M	3216	71.3 (0%)	196.2(4.3)	210.2(2.9)	219.3(2.3)	224.2(1.9)	187.7(12.8)	0.0(0.0)
BLACK	%	600	15.5 (1%)	8.4(1.3)	32.4(2.2)	18.7(1.3)	34.0(2.4)	6.5(1.5)	0.0(0.0)
	M	600	15.5 (1%)	179.5(10.8)	188.4(4.0)	199.5(5.3)	200.7(5.8)	184.5(9.4)	0.0(0.0)
HISPANIC	%	340	8.7 (1%)	25.3(2.8)	25.3(3.2)	16.7(2.2)	23.4(2.8)	9.2(1.4)	0.0(0.0)
	M	340	8.7 (1%)	181.8(4.4)	199.6(7.2)	207.2(7.9)	219.3(7.8)	165.7(12.4)	0.0(0.0)
OTHER	%	220	4.5 (1%)	5.8(1.8)	28.0(4.9)	15.3(2.3)	46.9(4.3)	4.0(1.3)	0.0(0.0)
	M	220	4.5 (1%)	166.9(16.6)	194.0(11.8)	206.1(7.3)	225.6(4.7)	222.6(76.2)	0.0(0.0)
PARENTAL EDUCATION									
LESS THAN HS	%	345	8.5 (6%)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
	M	345	8.5 (6%)	190.2(3.3)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
GRADUATED HS	%	1285	29.6 (4%)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
	M	1285	29.6 (4%)	0.0(0.0)	204.8(2.3)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
SOME EDU AFTER HS	%	826	18.5 (3%)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
	M	826	18.5 (3%)	0.0(0.0)	0.0(0.0)	215.1(2.3)	0.0(0.0)	0.0(0.0)	0.0(0.0)
GRADUATED COLLEGE	%	1807	40.5 (4%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)
	M	1807	40.5 (4%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	221.3(1.8)	0.0(0.0)	0.0(0.0)
UNKNOWN	%	113	3.0 (9%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)
	M	113	3.0 (9%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	180.7(5.8)	0.0(0.0)

% = PERCENTAGE; M = MEANPARTS TIMES 100

522

523

APPENDIX A

Consultants for the Development of the 1990 NAEP Objectives and Items

Appendix A

CONSULTANTS FOR THE DEVELOPMENT OF THE 1990 NAEP OBJECTIVES AND ITEMS

Joan Adler	West Windsor-Plainsboro High School Plainsboro, New Jersey
Valeria Allston	Vance Senior High School Henderson, North Carolina
Beverly Anderson	Education Commission of the States Denver, Colorado
Vincent Altamuro	Public School No. 22 Flushing, New York
Elaine Aoki	Seattle Public Schools Seattle Washington
Kathryn Au	Kamehemeha School Honolulu, Hawaii
Scott Baldwin	University of Miami Miami, Florida
Deborah Ball	Michigan State University East Lansing, Michigan
James Barafaldi	University of Texas at Austin Austin, Texas
Walter Barbe	HIGHLIGHTS FOR CHILDREN Honesdale, Pennsylvania
William Barstow	University of Georgia Athens, Georgia
Pat Bates	Timberlane Middle School Princeton, New Jersey
Jerry Bell	Harvard University Cambridge, Massachusetts

Richard Boyd	Mississippi Department of Education Jackson, Mississippi
George Bright	University of Houston Houston, Texas
Patricia Brinlee	South Carolina Department of Education Columbia, South Carolina
Bruce Brombacher	Jones Junior High School Westerville, Ohio
Linda Burroughs	Trenton State College Trenton, New Jersey
Thomas Butts	University of Texas Dallas, Texas
Rodger Bybee	Biological Sciences Curriculum Study Colorado Springs, Colorado
Carl Campion	Archmere Academy Wilmington, Delaware
Iris Carl	Houston Independent School District Houston, Texas
John Carroll	University of North Carolina Chapel Hill, North Carolina
Jackie Chalifoux	Hopewell Valley High School Hopewell, New Jersey
Jeanne Chall	Harvard Graduate School of Education Cambridge, Massachusetts
Audrey Champagne	American Association for the Advancement of Science Washington, DC
Carita Chapman	Chicago Public Schools Chicago, Illinois
Sam Chattin	Scottsburg Junior High School Scottsburg, Indiana

Peg Sibes	The William School New London, Connecticut
Michelle Commeyras	University of Illinois Champaign, Illinois
Diane Conradson	San Jose State College San Jose, California
Gerard Consuegra	Montgomery Public Schools Rockville, Maryland
Eric Cooper	Simon & Schuster School Group Morristown, New Jersey
Katherine Corley	Lansdown Middle School Baltimore, Maryland
George Cossman	University of Iowa Iowa City, Iowa
Debbie Coyle	Lawrence Intermediate School Lawrenceville, New Jersey
Jane Crowder	Pine Lake Middle School Issaquah, Washington
Donald Daugs	Utah State University Logan, Utah
Ed DeCrosta	Lawrence Intermediate School Lawrenceville, New Jersey
Paul DeHart Hurd	Stanford University Palo Alto, California
Franklin Demana	Ohio State University Columbus, Ohio
Joanne DeMaria	Smithsonian Institute Washington, DC
Kathy Donnellan	Springfield Public Schools Springfield, Massachusetts
John Dossey	Illinois State University Normal, Illinois

Floyd Downs	Arizona State University Tempe, Arizona
Dennis Doyle	Hudson Institute Alexandria, Virginia
Mary Dupis	Pennsylvania State University University Park, Pennsylvania
Peggy Dutcher	Michigan State Department of Education Lansing, Michigan
Joseph Exline	State Department of Education Richmond, Virginia
Alan Farstrup	International Reading Association Newark, Delaware
Robert Fisher	Illinois State University Normal, Illinois
Sam Forsythe	Loretto High School El Paso, Texas
Linda Foreman	Portland State University Portland, Oregon
Susan Fuhrman	Center for Policy Research in Education New Brunswick, New Jersey
Diane Gallo	West Windsor-Plainsboro Middle School Plainsboro, New Jersey
Robert Gardner	Salisbury School Salisbury, Connecticut
Victoria Purcell-Gates	University of Cincinnati Cincinnati, Ohio
Joyce Glatzer	Summit Public Schools Summit, New Jersey
Eunice Greer	Illinois State Board of Education Springfield, Illinois
James Guthrie	University of California at Berkeley Berkeley, California

John Guthrie	University of Maryland College Park, Maryland
Rosalina Hairston	National Association of Biology Teachers Reston, Virginia
Joan Hall	Summit County Day School Cincinnati, Ohio
Johnnie Hamilton	Thoreau Intermediate School Vienna, Virginia
Ted Hersey	St. Georges' School Newport, Rhode Island
Audrey Jackson	Parkway School District St. Louis, Missouri
Russ Jones	Sussex Elementary School Baltimore, Maryland
Kathleen Jongsma	Northside Independent School District San Antonio, Texas
Barbara Kapinus	Maryland State Department of Education Baltimore, Maryland
Jerome Kaplan	Seton Hall University Newark, New Jersey
Henry Kepner, Jr.	University of Wisconsin at Milwaukee Milwaukee, Wisconsin
Michael Kibby	State University of New York at Buffalo Buffalo, New York
JoEllen Killion	School District No. 12 North Glenn, Colorado
Jeremy Kilpatrick	University of Georgia Athens, Georgia
Charles Kish	Saratoga Junior High School Saratoga Springs, New York
Daniel Koretz	The Rand Corporation Washington, DC

Vicky Kouba	State University of New York at Albany Albany, New York
Carol Kubota	University of Washington Seattle, Washington
Paul Kuerbis	Biological Sciences Curriculum Study Colorado Springs, Colorado
Judith Langer	State University of New York at Albany Albany, New York
Douglas Lapp	National Sciences Resource Center Washington, DC
James Leitzel	Ohio State University Columbus, Ohio
Mary Lindquist	Columbus College Columbus, Georgia
Bobbi MacGregor	Franklin Township School Quakertown, New Jersey
Reynaldo Macias	University of Southern California Los Angeles, California
Carol Mazolla	Toll Gate Grammer School Pennington, New Jersey
Don Maxwell	Biological Sciences Curriculum Study Colorado Springs, Colorado
Joseph McInerney	Biological Sciences Curriculum Study Colorado Springs, Colorado
James Messersmith	John Witherspoon School Princeton, New Jersey
Vera Milz	Way School Bloomfield Hills, Michigan
Jim Minstrell	Mercer Island High School Mercer Island, Washington
Carol Mislan	Antheil Elementary School Ewing, New Jersey

Arlene Mitchell	University of Rhode Island Kingston, Rhode Island
Wendell Mohling	Shawnee Mission Northwest High School Shawnee Mission, Kansas
James Nelson	Harriton High School Rosemont, Pennsylvania
Rochelle Newman	West Windsor-Plainsboro Middle School Plainsboro, New Jersey
Floyd Nordland	Purdue University West Lafayette, Indiana
Joe Novak	Cornell University Ithaca, New York
Michael Padilla	University of Georgia Athens, Georgia
Scott Paris	University of Michigan Ann Arbor, Michigan
Jill Patterson	Oberlin High School Aurora, Colorado
Richard Patterson	Athens Academy Athens, Georgia
P. David Pearson	University of Illinois Champaign, Illinois
John Penick	University of Iowa Iowa City, Iowa
Charles Peters	Oakland Michigan Public Schools Pontiac, Michigan
Senta Raizen	The NETWORK, Inc. Washington, DC
Karen Reynolds	San Jose State College San Jose, California
William Robertson	Biological Sciences Curriculum Study Colorado Springs, Colorado

Stephanie Robinson	National Urban League New York, New York
Nancy Rosenberg	Great Neck, New York
Milli Rowen	Rocky Hill, New Jersey
Bud Rue	Lawrence High School Lawrenceville, New Jersey
Timothy Shanahan	University of Illinois at Chicago Circle Chicago, Illinois
Peggy O'Neill-Skinner	The Bush School Seattle, Washington
Karen Smith	Herrera Elementary School Phoenix, Arizona
Kathy Smith	Lawrence Intermediate School Lawrenceville, New Jersey
Kitty Lou Smith	Holms Intermediate School Alexandria, Virginia
Marilyn Smith	Beverley Vista Schools Beverley Hills, California
Robert Smith	Michigan State Department of Education Lansing, Michigan
Anne Stallman	University of Illinois Champaign, Illinois
Robert Swartz	University of Massachusetts Boston, Massachusetts
Bonita Talbot-Wylie	Excelsior Elementary School Excelsior, Minnesota
Juliana Texley	Richmond High School Richmond, Michigan
Thomas Tucker	Colgate University Hamilton, New York

Anna Ulrich	New Mexico State Department of Education Santa Fe, New Mexico
Sheila Valencia	University of Washington Seattle, Washington
Bert Waits	Ohio State University Columbus, Ohio
Herbert Walberg	University of Illinois at Chicago Circle Chicago, Illinois
Constance Weaver	Western Michigan University Portage, Michigan
David Webster	The Dexter School Brookline, Massachusetts
Gerald Wheeler	Montana State University Bozeman, Montana
John Wills	National Governors' Association Washington, DC
Peter Winograd	University of Kentucky Lexington, Kentucky
Karen Wixson	University of Michigan Ann Arbor, Michigan
Linda Zimmerer	California Assessment Program Sacramento, California
John Zorri	John Witherspoon Middle School Princeton, New Jersey

APPENDIX B

Reporting Subgroups and Derived Variables Used for Reporting and Conditioning

NAEP 1990 REPORTING SUBGROUPS

DSEX (Gender)

The variable SEX on the student files is the gender of the student being assessed, as taken from school records. For a few students, data for this variable was missing and was imputed by ETS after the assessment. The resulting variable DSEX on the student file contains a value for every student and should be used for gender comparisons among students.

DRACE (Race/ethnicity)

The variable DRACE on the student file is an imputed definition of race/ethnicity, derived from up to three sources of information. This variable is used for race/ethnicity subgroup comparisons within the 1990 assessment and among the 1990, 1988, 1986, and 1984 assessments. Two items from the student demographics questionnaire were used in the determination of derived race/ethnicity:

Demographic Item Number 2:

2. If you are Hispanic, what is your Hispanic background?

- ⬭ I am not Hispanic.
- ⬭ Mexican, Mexican American, or Chicano
- ⬭ Puerto Rican
- ⬭ Cuban
- ⬭ Other Spanish or Hispanic background

Students who responded to item number 2 by filling in the second, third, fourth, or fifth oval were considered Hispanic. For students who filled in the first oval, did not respond to the item, or provided information that was illegible or could not be classified, responses to item number 1 were examined in an effort to determine race/ethnicity. Item number 1 read as follows:

Demographic Item Number 1:

1. Which best describes you?

- ⬭ White (not Hispanic)
- ⬭ Black (not Hispanic)

- Hispanic ("Hispanic" means someone who is Mexican, Mexican American, Chicano, Puerto Rican, Cuban, or from some other Spanish or Hispanic background.)

- Asian or Pacific Islander ("Asian or Pacific Islander" means someone who is Chinese, Japanese, Korean, Filipino, Vietnamese, or from some other Asian or Pacific Island background.)

- American Indian or Alaskan Native ("American Indian or Alaskan Native" means someone who is from one of the American Indian tribes, or one of the original people of Alaska.)

- Other (What?) ________

Students' race/ethnicity was then assigned to correspond with their selection. For students who filled in the sixth oval ("Other"), provided illegible information or information that could not be classified, or did not respond at all, observed race/ethnicity (RACE on the data files), if provided from school records, was used.

Derived race/ethnicity could not be determined for students who did not respond to background items 1 or 2 and for whom an observed race/ethnicity was not provided.

RACE (Observed Race/ethnicity)

The variable RACE on the student files is the race/ethnicity of the student being assessed, as observed and recorded by the exercise administrator. Observed race/ethnicity was used in NAEP assessments before 1984. This variable is used for race/ethnicity subgroup comparisons to pre-1984 assessments.

TOC, SCTOC (Type of community)

NAEP assigned each participating school to one of four type of categories designed to provide information about the communities in which the schools are located. These categories are contained on the student data files as the variable TOC and on the school files as SCTOC.

The type of community categories consist of three "extreme" types of communities and one "other" type of community. Schools were placed into these categories on the basis of information about the type of community, the size of its population (as of the 1980 Census), and an occupational profile of residents provided by school principals before the assessment. The principals completed estimates of the percentage of students whose parents fit into each of six occupational categories. The type of community categories are as follows:

1 - Extreme Rural: Students in this group live outside metropolitan statistical areas, live in areas with a population below 10,000, and attend schools where many of the students' parents are farmers or farm workers.

2 - Disadvantaged Urban: Students in this group live in metropolitan statistical areas and attend schools where a high proportion of the students' parents are on welfare or are not regularly employed.

3 - Advantaged Urban: Students in this group live in metropolitan statistical areas and attend schools where a high proportion of the students' parents are in professional or managerial positions.

4 - Other: Students in this category attend schools in areas other than those defined as advantaged urban, disadvantaged urban, or extreme rural.

PARED (Parents' education level)

The variable PARED on the student file is derived from responses to two questions, B003501 and B003601, in the student demographic questionnaire. Students were asked to indicate the extent of their mother's education (B003501) by choosing one of the following:

- She did not finish high school.
- She graduated from high school.
- She had some education after high school.
- She graduated from college.
- I don't know.

Students were asked to provide the same information about the extent of their father's education (B003601) by choosing one of the following:

- He did not finish high school.
- He graduated from high school.
- He had some education after high school.
- He graduated from college.
- I don't know.

The information was combined into one parental education reporting category (PARED) as follows: If a student indicated the extent of education for only one parent, that level was included in the data. If a student indicated the extent of education for both parents, the higher of the two levels was included in the data. For students who did not know the level of education for both parents or did not know the level of education for one parent and did not respond for the other, the parental education level was classified as unknown. If the student did not respond for both parents, the student was recorded as having provided no response.

REGION, SREGION (Region of the country)

In addition to overall responses, NAEP computed data for four geographical regions in the United States. States were assigned to regions as follows:

NORTHEAST	SOUTHEAST	CENTRAL	WEST
Connecticut	Alabama	Illinois	Alaska
Delaware	Arkansas	Indiana	Arizona
District of Columbia	Florida	Iowa	California
Maine	Georgia	Kansas	Colorado
Maryland	Kentucky	Michigan	Hawaii
Massachusetts	Louisiana	Minnesota	Idaho
New Hampshire	Mississippi	Missouri	Montana
New Jersey	North Carolina	Nebraska	Nevada
New York	South Carolina	North Dakota	New Mexico
Pennsylvania	Tennessee	Ohio	Oklahoma
Rhode Island	Virginia*	South Dakota	Oregon
Vermont	West Virginia	Wisconsin	Texas
Virginia*			Utah
			Washington
			Wyoming

* The part of Virginia that is included in the Washington, DC, metropolitan statistical area is included in the Northeast region; the remainder of the state is included in the Southeast region.

DGRADE, MODGRD (Grade in school)

To enhance the usefulness of the data, in 1984 NAEP began sampling students by grade as well as by age. The ages sampled in assessments since 1984—9, 13, and 17—match the ages sampled in earlier assessments. However, some of the modal grades (the grade attended by most students of a particular age) for the ages sampled have varied in the last three assessments because of changes in how student age was determined and changes in the times of the year that students were tested.

In the 1990 main assessment, the respective modal grades for ages 9, 13, and 17 are 4, 8, and 12. Student age for all three cohorts was determined on a calendar-year basis; all students were tested at the same times of the year. The 1990 bridge samples, by definition designed to match previous assessment characteristics, sampled varying student cohorts.

The 1990 main sample included many students in each cohort who were both age-eligible (age 9, 13, or 17) and grade-eligible (attending respectively grade 4, 8, or 12). However, because NAEP collected data by grade or age, each cohort also included students who were age-eligible but not in the modal grade, and students who were grade-eligible but not of the modal age (the age of most students attending the particular grade).

For each 1990 sample, results for students in a particular grade can be selected using (1) the variable DGRADE, the student's actual grade at time of testing, on the student file, or (2) the student file variable MODGRD (setting MODGRD to a value of 2 will select those students who are in the modal grade).

DAGE, MODAGE (Student age)

Results for students at a particular age can be selected using (1) the student file variable DAGE, the student's age as of December 31, 1989 or (2) the student file variable MODAGE. Because NAEP collected data by grade *or* age, each main sample student cohort includes students who were both age-eligible and grade-eligible, students who were age-eligible but not in the modal grade, and students who were grade-eligible but not of the modal age. The main assessment modal age (the age of most of the students in the grade sample) is age 9 for fourth graders, age 13 for eighth graders, and age 17 for twelfth graders. A value of 1 for MODAGE indicates that the student is younger than the modal age; a value of 2 indicates that the student at the modal age; a value of 3 indicates that the student is older than the modal age.

DERIVED VARIABLES USED FOR REPORTING AND CONDITIONING

Several NAEP variables on the student data files were formed for use in reporting and conditioning from the systematic combination of response values for one or more items from either the student demographic questionnaire, the student subject-area background questionnaires, the teacher questionnaire, or the school questionnaire.

These variables maximize use of the data, incorporate a larger segment of the population, and save analysis costs by grouping items that measure similar characteristics into one variable. The derivation of each of these variables is explained below.

Variables Derived from the Student Demographic Questionnaires

HOMEEN2 (Home environment—Articles [of 4] in the home)

The variable HOMEEN2 was created from the responses to student demographic items B000901, B000903, B000904, and B000905 concerning articles found in the student's home (newspaper, encyclopedia, more than 25 books, and magazines). The values for this variable were derived as follows:

1 0-2 types	The student responded to at least two items and answered YES to two or fewer.
2 3 types	The student answered YES to three items.
3 4 types	The student answered YES to four items.
8 Omitted	The student answered fewer than two items.

PARWK4 (Economic Support—Which parents work)

For age class 9, PARWK4 was created from responses to items B005901 and B006101, which asked if the student's mother (or stepmother) and father (or stepfather) worked for pay. The values for PARWK4 were derived as follows:

1 Both mother & father	The student answered YES to both B005901 and B006101.
2 Father only	The student answered NO or DON'T LIVE WITH to B005901 and YES to B006101.
3 Mother only	The student answered NO or DON'T LIVE WITH to B006101 and YES to B005901.
4 Something else	Any other combination of responses

8 No response	The student did not respond to one or both items.
9 Mult.	The student filled in more than one oval for both items

NCOMP (Number of computer science courses taken)

For age class 17, NCOMP was created from responses to items B005312 and B005313 concerning the student's coursework in computer science. The values for NCOMP were derived as follows:

1 0	The student answered HAVE NOT to both courses.
2 1	The student answered HAVE to one course.
3 2	The student answered YES to both courses.
8 No response	The student did not respond to one or both items.
9 Mult. & out-of-range	The student filled in more than one oval for both items.

NMATH (Highest level of mathematics courses taken)

For age class 17, NMATH was created from responses to items B005301 through B005307 concerning the student's coursework in mathematics. The values for NMATH were derived as follows:

1 Gen. math or pre-algebra	The student answered HAVE NOT to all items or HAVE to B005301 or B005302 and HAVE NOT to all others.
2 Algebra	The student answered HAVE to B005303 and HAVE NOT to B005304, B005305, B005306, and B005307.
3 Geometry	The student answered HAVE to both B005303 and B005305 and HAVE NOT to B005304, B005306, and B005307.
4 Algebra 2	The student answered HAVE to B005304 or B005306 but HAVE NOT to B005307.
5 Calculus	The student answered HAVE to B005307.
6 Something else	Any other response combination
8 No response	The student did not respond to any item.

NSCI (Highest level of science courses taken)

For age class 17, NSCI was created from responses to items B005308 through B005311, which concerned the student's coursework in science. The values for NSCI were derived as follows:

1 No biology	The student answered HAVE NOT to all items or HAVE to B005308 and other than HAVE to all other items.
2 Biology	The student answered HAVE to B005309 and other than HAVE to both B005310 and B005311.
3 Chemistry	The student answered HAVE to both B005309 and B005310 and other than HAVE to B005311.
4 Physics	The student answered HAVE to B005309, B005310, and B005311.
5 Something else	Any other response combination
8 No response	The student answered none of the items.

SINGLEP (How many parents live at home)

SINGLEP was created from items B005601 and B005701, which asked whether the student's mother (or stepmother) and father (or stepfather) lived at home with the student. The values for SINGLEP were derived as follows:

1 2 parents at home	The student answered YES to both items.
2 1 parent at home	The student answered YES to B005601 and NO to B005701, or YES to B005701 and NO to B005601.
3 Neither at home	The student answered NO to both items.
8 Omitted	The student did not respond to or filled in more than one oval for one or both items.

Variables Derived from the Mathematics Background Questionnaires

PERCMAT (Students' perception of mathematics, age 13/grade 8 and age 17/grade 12)

PERCMAT was created from items M810701 through M810705 in the mathematics background questionnaire, which asked students about their perceptions of each of five statements:

M810701 I like mathematics.
M810702 Almost all people use mathematics in their jobs.
M810703 I am good in mathematics.
M810704 Mathematics is more for boys than for girls.
M810705 Mathematics is useful for solving everyday problems.

For each item, the student could respond as follows:

1. Strongly agree
2. Agree
3. Undecided
4. Disagree
5. Strongly disagree

To derive PERCMAT, first the values for one item (M810704) were reversed (e.g., "strongly disagree" became 1). Then, for each of the five items, values 3, 4, and 5 were combined to create one value (new value 3). PERCMAT was determined by adding the values for the five items and dividing by five to obtain a mean. The mean was then recoded as follows:

1 - 1.67 = 1 Strongly agree
1.68 - 2.33 = 2 Agree
2.34 - 3 = 3 Undecided, disagree, or strongly disagree

The student had to answer at least one of the five items to get a value for PERCMAT.

PERCMA2 (Students' perception of mathematics, age 9/grade 4)

PERCMA2 was created from items M811101 through M8111105 in the mathematics background questionnaire, which asked students about their perceptions of each of five statements:

M811101 I like mathematics.
M811102 Almost all people use mathematics in their jobs.
M811103 I am good in mathematics.
M811104 Mathematics is more for boys than for girls.
M811105 Mathematics is useful for solving everyday problems.

For each item, the student could respond as follows:

1. Agree
2. Undecided
3. Disagree

To derive PERCMA2, first the values for one item (M811104) were reversed ("disagree" became 1 and "agree" became 3). Then, for each of the five items, values 2 and 3 were combined to create one value (new value 2). PERCMA2 was determined by adding the values for the five items and dividing by five to obtain a mean. The mean was then recoded as follows:

1 - 1.50 = 1 Agree
1.51 - 2 = 2 Undecided or disagree

ALGCALC (Algebra and calculus course taking, age 17/grade 12)

ALGCALC was created from five of the items in the mathematics background questionnaire that asked students how long they had taken certain mathematics courses:

M811003	Introduction to algebra or pre-algebra
M811004	First-year algebra
M811006	Second-year algebra
M811008	Pre-calculus, third-year algebra, elementary functions, or analysis
M811011	Calculus

For each item, the student could respond as follows:

1. More than 1 year
2. 1 school year
3. ½ year or less
4. Not studied

The values for ALGCALC were derived as follows:

Not studied	The student did not answer MORE THAN 1 YEAR or 1 SCHOOL YEAR to M811003
Pre-algebra	The student answered MORE THAN 1 YEAR or 1 SCHOOL YEAR to M811003 but not to M811004
1st year algebra	The student answered MORE THAN 1 YEAR or 1 SCHOOL YEAR to M811004 but not to M811006
2nd year algebra	The student answered MORE THAN 1 YEAR or 1 SCHOOL YEAR to M811006 but not to M811008
3rd year algebra	The student answered MORE THAN 1 YEAR or 1 SCHOOL YEAR to M811008 but not to M811011
Calculus	The student answered MORE THAN 1 YEAR or 1 SCHOOL YEAR to M811011

GEOTRIG (Geometry and trigonometry course taking, age 17/grade 12)

GEOTRIG was created from two of the items in the mathematics background questionnaire that asked students how long they had taken certain mathematics courses:

M811005 Geometry
M811007 Trigonometry

For each item, the student could respond as follows:

1. More than 1 year
2. 1 school year
3. ½ year or less
4. Not studied

The values for GEOTRIG were derived as follows:

Not studied	The student did not answer MORE THAN 1 YEAR or 1 SCHOOL YEAR to M811005
Geometry	The student answered MORE THAN 1 YEAR or 1 SCHOOL YEAR to M811005 but not to M811007
Trigonometry	The student answered MORE THAN 1 YEAR or 1 SCHOOL YEAR to M811007

Variables Derived from the Mathematics Teacher Questionnaire

TCERTIF (Type of teaching certificate)

Items T030301 through T030305 in the mathematics teacher questionnaire were combined to produce TCERTIF. The following rules were used to determine the three values for TCERTIF.

1 Mathematics	The teacher responded YES to either T030303 or T030304
2 Education	The teacher responded YES to either T030301 or T030302 and NO to T030303 and T030304
3 Else	Any other response

TUNDMAJ (Undergraduate major)

Items T023301, T023311, T023307, and T023313 in the mathematics teacher questionnaire were used to determine TUNDMAJ as follows:

1 Mathematics	The teacher responded YES to T023311
2 Education	The teacher responded YES to T023301 and NO to T023311

3 Else The teacher responded YES to T023307 or T023313 and NO to T023311 and T023301

TGRDMAJ (Graduate major)

Items T023401, T023411, T023407, and T023413 in the mathematics teacher questionnaire were used to determine TUNDMAJ as follows:

1 Mathematics The teacher responded YES to T023411

2 Education The teacher responded YES to T023401 and NO to T023411

3 Else The teacher responded YES to T023407 or T023413 and NO to T023411 and T023401

TMATCRS (Number of mathematics areas in which courses were taken)

TMATCRS was derived from items T030407 through T030411, T030413, and T030414 in the mathematics teacher questionnaire. Those items asked how many courses the teacher had taken in a variety of areas. TMATCRS was derived by obtaining a count of the number of times (of seven) that the teacher responded to the number-of-courses category "1," "2," or "3 or more". The levels of TMATCRS were then defined as:

1 0 to 3 courses
2 4 to 5 courses
3 6 to 7 courses

The teacher had to answer at least one of these items to receive a value for TMATCRS.

TEMPHNO (Teacher's emphasis in numbers and operations)

TEMPHNO was derived from mathematics teacher questionnaire items T031501, T031502, T031503, T031515, and T031516. The variable was derived by first combining categories three (little emphasis) and four (none) for each item and changing the value for that category to three. The mean of the values for all five items was then recoded as follows:

1 - 1.67	1	Heavy emphasis
1.68 - 2.33	2	Moderate emphasis
2.34 - 3	3	Little or no emphasis

The teacher had to answer at least one of these items to receive a value for TEMPHNO.

TEMPHPS (Teacher's emphasis in data analysis, probability, and statistics)

TEMPHPS was derived from mathematics teacher questionnaire items T031506 and T031507. The variable was derived by first combining category three (little emphasis) and four (none) for both items and changing the value for that category to three. The mean of the values for both items was recoded as follows:

1 - 1.67	1	Heavy emphasis
1.68 - 2.33	2	Moderate emphasis
2.34 - 3	3	Little or no emphasis

The teacher had to answer at least one of the items to receive a value for TEMPHPS.

Variables Derived from the Science Background Questionnaire

NEXPER (Engagement in science-related experiments)

For grades 4, 8, and 12, NEXPER was created from items K810101 to K810106, which asked if students had performed experiments related to science. NEXPER was derived by obtaining a count of the number of times (out of six) that the student responded "Yes". Then the levels of NEXPER were defined as follows:

1 Performed no experiments
2 Performed 1 or 2 experiments
3 Performed 3 or 4 experiments
4 Performed 5 or 6 experiments
M Missing for all items

SCIACT (Engagement in science-related activities)

For grades 4, 8, 12, SCIACT was created from items K810602 to K810605, which asked if students had done activities related to science. SCIACT was derived by counting the number of times (out of four) that the student responded "Almost every day," "Several times a week," or "About once a week". Then the levels of SCIACT were defined as follows:

1 Did no activities on at least a weekly bases
2 Did 1 or 2 activities on at least a weekly bases
3 Did 3 or 4 activities on at least a weekly bases
M Missing for all items

SCIACT2 (Engagement in science-related activities - 2nd set of activities)

For grades 8 and 12, SCIACT2 was created from items S402003, S402006, S402007, and S402008, which asked if student had done activities related to science that were requested by the

teacher. SCIACT2 was derived by counting the number of times (out of four) that the student responded "Almost every day," "Several times a week," or "About once a week". Then the levels of SCIACT2 were defined as follows:

1 Did no activities on at least a weekly bases
2 Did 1 or 2 activities on at least a weekly bases
3 Did 3 or 4 activities on at least a weekly bases
M Missing for all items

Variables Derived from the Science Teacher Questionnaire

TYREXP1 (Number of years teaching elementary or secondary education)

TYREXP1 was created from item T030001, which asked teachers the total number of years taught. TYREXP1 was derived by grouping years of experience together. The levels of TYREXP1 were defined as follows:

1 5 years experience or less
2 6 to 10 years experience
3 11 years experience or more

TYREXP2 (Number of years teaching elementary or secondary education)

TYREXP2 was created from item T030001, which asked teachers the total number of years taught. TYREXP2 was derived by grouping years of experience together. The levels of TYREXP2 were defined as follows:

1 10 years experience or less
2 11 to 24 years experience
3 25 years experience or more

TYRSCI1 (Number of years teaching science)

TYRSCI1 was created from item T034701, which asked teachers the total number of years teaching science. TYRSCI1 was derived by grouping years of teaching science together. The levels of TYRSCI1 were defined as follows:

1 5 years teaching science or less
2 6 to 10 years teaching science
3 11 years teaching science or more

TYRSCI2 (Number of years teaching science)

TYRSCI2 was created from item T034701, which asked teachers the total number of years teaching science. TYRSCI2 was derived by grouping years of teaching science together. The levels of TYRSCI2 were defined as follows:

1 10 years teaching science or less
2 11 to 24 years teaching science
3 25 years teaching science or more

TCERT1 (Teacher certification in science)

TCERT1 was created from items T030306 to T030308, which asked if teachers were certified in various areas. TCERT1 was derived by determining whether or not teachers were certified in science and assigning the following codes and labels:

Certified in science	1 Yes
Not certified in science	2 No

TCERT2 (Teacher certification in science)

TCERT2 was created from items T030306 to T030308, which asked if teachers were certified in various areas. TCERT2 was derived by determining the level at which teachers were certified to teach science and assigning the following codes and labels:

Certified in elementary science	1 Elementary science
Certified in any secondary science	2 Secondary science
Not certified in science	3 Not certified

TCRSBIO (College biology courses completed)

TCRSBIO was created from items T032308 to T032316, which asked if teachers had completed college courses related to biology. TCRSBIO was derived by counting the number of times (out of nine) that a teacher responded "One," "Two," or "Three or more". Then the levels of TCRSBIO were defined as follows:

1 Took courses in 2 or fewer subjects related to biology
2 Took courses in 3 to 5 subjects related to biology
3 Took courses in 6 or more subjects related to biology
M Missing for all items

TCRSCHM (College chemistry courses completed)

TCRSCHM was created from items T032317 to T032321, which asked if teachers had completed college courses related to chemistry. TCRSCHM was derived by counting the number of times (out of five) that a teacher responded "One," "Two," or "Three or more". Then the levels of TCRSCHM were defined as follows:

1 Took no courses in subjects related to chemistry
2 Took courses in 1 or 2 subjects related to chemistry
3 Took courses in 3 or more subjects related to chemistry
M Missing for all items

TCRSPHY (College physics courses completed)

TCRSPHY was created from items T032322 to T032328, which asked if teachers had completed college courses related to physics. TCRSPHY was derived by counting the number of times (out of seven) that a teacher responded "One," "Two," or "Three or more". Then the levels of TCRSPHY were defined as follows:

1 Took no courses in subjects related to physics
2 Took courses in 1 or 2 subjects related to physics
3 Took courses in 3 or more subjects related to physics 3
M Missing for all items

TCRSES (College earth science courses completed)

TCRSES was created from items T032329 to T032335, which asked if teachers had completed college courses related to earth science. TCRSES was derived by counting the number of times (out of seven) that a teacher responded "One," "Two," or "Three or more". Then the levels of TCRSES were defined as follows:

1 Took no courses in subjects related to earth science
2 Took courses in 1 or 2 subjects related to earth science
3 Took courses in 3 or more subjects related to earth science
M Missing for all items

TPREPSC (Preparedness to teach science topics)

TPREPSC was created from items T033301 to T033319, which asked if teachers felt prepared to teach various science subjects. Each of the items were given a value of 3 if the teacher responded "Well prepared," 2 if the teacher responded "Somewhat prepared," 1 if the teacher responded "Ill prepared," and 0 if the teacher did not respond. TPREPSC was derived by taking the average of the 19 items and assigning the following codes and labels:

2.34 or greater	1	Well prepared
1.34-2.33	2	Somewhat prepared
1.33 or less	3	Ill prepared
	M	Missing for all items

TPREPLS (Preparedness to teach life science topics)

TPREPLS was created from items T033301 to T033308, which asked if teachers felt prepared to teach various life science subjects. Each of the items were given a value of 3 if the teacher responded "Well prepared," 2 if the teacher responded "Somewhat prepared," 1 if the teacher responded "Ill prepared," and 0 if the teacher did not respond. TPREPLS was derived by taking the average of the eight items and assigning the following codes and labels:

2.34 or greater	1	Well prepared
1.34-2.33	2	Somewhat prepared
1.33 or less	3	Ill prepared
	M	Missing for all items

TPREPPS (Preparedness to teach physical science topics)

TPREPPS was created from items T033309 to T033314, which asked if teachers felt prepared to teach various physical science subjects. Each of the items were given a value of 3 if the teacher responded "Well prepared," 2 if the teacher responded "Somewhat prepared," 1 if the teacher responded "Ill prepared," and 0 if the teacher did not respond. TPREPLS was derived by taking the average of the eight items and assigning the following codes and labels:

2.34 or greater	1	Well prepared
1.34-2.33	2	Somewhat prepared
1.33 or less	3	Ill prepared
	M	Missing for all items

TPREPES (Preparedness to teach earth science topics)

TPREPES was created from items T033315 to T033319, which asked if teachers felt prepared to teach various earth science subjects. Each of the items were given a value of 3 if the teacher responded "Well prepared," 2 if the teacher responded "Somewhat prepared," 1 if the teacher responded "Ill prepared," and 0 if the teacher did not respond. TPREPES was derived by taking the average of the five items and assigning the following codes and labels:

2.34 or greater	1	Well prepared
1.34-2.33	2	Somewhat prepared
1.33 or less	3	Ill prepared
	M	Missing for all items

TEMPHLS (Emphasis on life science topics)

TEMPHLS was created from items T034001 to T034008, which asked if teachers emphasized various life science topics. Each of the items were given a value of 3 if the teacher responded "Heavy emphasis," 2 if the teacher responded "Moderate emphasis," 1 if the teacher responded "Little emphasis," or 0 if the teacher responded "None" or did not respond. TEMPHLS was derived by taking the average of the eight items and assigning the following codes and labels:

2.34 or greater	1	Heavy emphasis
1.34-2.33	2	Moderate emphasis
1.33 or less	3	Little or no emphasis
	M	Missing for all items

TEMPHPH (Emphasis on physical science topics)

TEMPHPH was created from items T034009 to T034014, which asked if teachers emphasized various physical science topics. Each of the items were given a value of 3 if the teacher responded "Heavy emphasis," 2 if the teacher responded "Moderate emphasis," 1 if the teacher responded "Little emphasis," or 0 if the teacher responded "None" or did not respond. TEMPHPH was derived by taking the average of the six items and assigning the following codes and labels:

2.34 or greater	1	Heavy emphasis
1.34-2.33	2	Moderate emphasis
1.33 or less	3	Little or no emphasis
	M	Missing for all items

TEMPHES (Emphasis on earth science topics)

TEMPHES was created from items T034015 to T034019, which asked if teachers emphasized various physical science topics. Each of the items were given a value of 3 if the teacher responded "Heavy emphasis," 2 if the teacher responded "Moderate emphasis," 1 if the teacher responded "Little emphasis," or 0 if the teacher responded "None" or did not respond. TEMPHES was derived by taking the average of the five items and assigning the following codes and labels:

2.34 or greater	1	Heavy emphasis
1.34-2.33	2	Moderate emphasis
1.33 or less	3	Little or no emphasis
	M	Missing for all items

TSCIACT (Engagement in science-related activities)

TSCIACT was created from items T034105, T034106, T034107, and T034109, which asked teachers how often students in their class did various activities related to science. TSCIACT was derived by counting the number of times (out of four) that the teacher responded "Almost every day," "Several times a week," or "About once a week". Then the levels of TSCIACT were defined as follows:

1 Did no activities on at least a weekly bases
2 Did 1 or 2 activities on at least a weekly bases
3 Did 3 or 4 activities on at least a weekly bases
M Missing for all items

Variables Derived from the School Questionnaire

SPOLICY (Changes in school policy since 1984-85)

School questionnaire items C028101 to C028103 and C028105 to C028109 were used to derive the variable SPOLICY. Those items asked if changes had been made in school policy in a variety of areas. SPOLICY was derived by obtaining a count of the number of times (of eight) that the response was YES to these items. The levels of SPOLICY were then defined as:

1 0 to 2 changes
2 3 to 4 changes
3 5 to 8 changes

SPROBS (Problems in the school)

School questionnaire items C028201 through C028211 were used to derive the variable SPROBS. Those items asked if problems existed in the school in a variety of areas. To derive SPROBS, category one (serious) and two (moderate) for each item were combined into a new category one. Category three was recoded as category two and category four was recoded as category three. The mean of the values for all 11 items were then recoded as follows:

1 - 1.67 = 1 Moderate to serious
1.68 - 2.33 = 2 Minor
2.34 - 3 = 3 Not a problem

PCLUNCH (Percent in school lunch program)

The values for the variable PCLUNCH on the student data files were calculated from the school questionnaire variables C025010 (number of students in subsidized lunch program) and C026202 (total enrollment as of October 1, 1989). The value for C025010 was divided by the value for C026202 to create the value for PCLUNCH.

SMATOFF (School offers pre-calculus and calculus, age 17/grade 12)

SMATOFF was created from two school questionnaire items asking whether pre-calculus (C030410) and calculus (C030411) were taught in the school. The values of SMATOFF were calculated as follows:

Both	School answered YES to both C030410 and C030411
Only calculus	School answered YES to C030411 and NO to C030410
Pre-calculus	School answered YES to C030410 and NO to C030411
Neither	School answered NO to both items

Variables Derived from Mathematics Items

CALCUSE (Calculator-usage index)

CALCUSE was created from noncognitive questions included in mathematics blocks M8 and M9. Students were provided a scientific calculator to use in answering the cognitive questions in those two blocks. Each cognitive item was followed by the question *"Did you use a calculator on this question?"*. The responses to these questions were used to derive the variable CALCUSE.

The cognitive items in blocks M8 (18 items) and M9 (20 items) were classified into one of three categories -- calculator-active, calculator-inactive, and calculator-neutral. Calculator-active items required the use of a calculator for their solution. Calculator-inactive items asked questions for which the use of a calculator was inappropriate. Calculator-neutral items could be solved with or without a calculator. The category for each of the calculator items is identified in column 109 of the machine-readable catalog files for the student data (1 = calculator-active, 2 = calculator-inactive, 3 = calculator-neutral).

Block M8 contained three calculator-active items, seven calculator-inactive items, and eight calculator-neutral items. Block M9 contained five calculator-active items, ten calculator-inactive items, and five calculator-neutral items. Blocks M8 and M9 each appeared in a total of three test booklets. However, one booklet contained both blocks M8 and M9. Therefore, at least one block of calculator items appeared in five of the seven assessment booklets.

The calculator-usage index for students assigned a booklet containing only block M8 was based on 10 items; the index for students assigned a booklet containing only block M9 was based on 15 items; and the index for students assigned a booklet containing both blocks M8 and M9 was based on 25 items.

CALCUSE had two levels, defined as follows:

1 High Students who used the calculator appropriately (i.e., used it for the calculator-active items and did not use it for the calculator-inactive items) at least 85 percent of the time and indicated they had used the calculator for at least half of the calculator-active items they were presented.

2 Other Students who did not use the calculator appropriately at least 85 percent of the time or indicated that they had used the calculator for less than half of the calculator-active items they were presented.

The percentage of appropriate calculator usage was determined using only those items that were answered by the student. Omitted items were not included as part of the denominator in calculating the percentage of appropriate calculator use.

Other Derived Variables

NUMCOR (Number correct within booklet)
PCTCOR (Percent correct within booklet)
LOGITP (Logit percent correct within booklet)
ZSCORE (Standardized logit percent correct within booklet)

The student file variables NUMCOR, PCTCOR, and LOGITP are statistics describing a student's responses to the main assessment cognitive items in the booklet he or she received. (Note: Each student was administered one of seven different assessment booklets, each of which contained a different combination of items from the total item pool.) These three variables were used to create a standardized logit score, ZSCORE.

NUMCOR is the number of correct responses a student made to the items in the booklet; PCTCOR is the percent of correct responses, calculated as the number of correct responses (NUMCOR) divided by the total number of items in the booklet. If NUMCOR equaled zero, PCTCOR was set to .0001; if NUMCOR equaled the total number of items in the booklet, PCTCOR was set to .9999.

A logit score, LOGITP, was calculated for each student by the following formula:

$$\text{LOGITP} = \ln\left[\frac{PCTCOR}{1 - PCTCOR}\right]$$

LOGITP was then restricted to a value x, such that $-3 \leq x \leq 3$. After computing LOGITP for each student, the mean and standard deviation was calculated for each booklet as the first step in standardizing the logit scores. The standardized logit score, ZSCORE, was then calculated for each student by the following formula:

$$\text{ZSCORE} = \left[\frac{LOGITP - mean\ logit}{standard\ deviation}\right]$$

SCHMATH (School-level mathematics mean logit score)
SCHRDG (School-level reading mean logit score)
SCHSCI (School-level science mean logit score)

These school-level mean proficiency variables were used in conditioning procedures to take into account differences in school proficiency. For each booklet, weighted frequency distributions were obtained of the number of correct responses for the students taking that booklet. A percentile rank for each student was determined from the frequency distribution of the booklet that student received. The logit of the percentile rank was calculated as:

$$\ln\left[\frac{\textit{percentile rank}}{1 - \textit{percentile rank}}\right] .$$

For each school, the weighted mean of the logits for the students in that school was calculated. Each student was then assigned that mean as his or her school-level mean logit score value.

APPENDIX C

Distribution of Weight Components for 1990 NAEP Samples

Appendix C

DISTRIBUTION OF WEIGHT COMPONENTS FOR 1990 NAEP SAMPLES

The following tables, which are cited and described throughout Chapter 10, "Weighting Procedures and Estimation of Sampling Variance," show the distribution of student and excluded student weight components for the 1990 NAEP samples, including base weights, the various nonresponse adjustment factors, trimming factors, and poststratification factors.

Table C-1

Distribution of Student Base Weights, Winter Main Samples

Sample	Number of Cases	Mean	Standard Deviation	Minimum	25th Percentile	Median (50%)	75th Percentile	Maximum
AGE 9/GRADE 4 WINTER MAIN SAMPLES								
Booklet 28	1733.	2402.4	1132.7	1107.9	1751.8	1821.6	3556.5	6181.8
Booklets 1-10	6192.	384.6	226.5	109.5	238.8	293.1	517.9	1496.5
Booklet 11-17	4573.	925.9	580.1	264.3	546.0	692.6	1160.0	3369.5
Booklets 18-24	4320.	386.9	224.5	109.5	240.0	293.1	522.7	1496.5
Excluded students	1282.	238.8	109.1	91.2	179.5	180.9	325.6	947.8
AGE 13/GRADE 8 WINTER MAIN SAMPLES								
Booklet 25	1778.	2121.7	1499.7	869.9	2.0	2.0	2.2	2.9
Booklets 1-7	4732.	388.5	212.4	113.5	250.1	319.6	516.3	2464.6
Booklets 15-21	4764.	391.5	216.7	113.5	250.1	319.6	516.3	2464.6
Booklets 8-14	4624.	769.8	439.9	308.0	477.0	581.3	1005.9	5103.6
Excluded students	1129.	229.9	135.3	99.6	162.6	175.3	228.2	1512.1
AGE 17/GRADE 12 WINTER MAIN SAMPLES								
Booklet 25	1649.	1586.5	813.7	364.3	947.0	1424.5	2048.7	3938.8
Booklets 1-7	4367.	298.9	159.7	40.0	193.1	269.6	370.6	1038.4
Booklets 15-21	4408.	297.1	157.6	40.0	192.3	269.6	370.6	1038.4
Booklets 8-14	4220.	604.4	351.3	121.4	348.6	529.6	735.1	1969.4
Excluded students	807.	188.3	2.1	2.6	4.3	5.7	6.8	16.9

Table C-2

Distribution of Student Base Weights, Spring Main Samples

Sample	Number of Cases	Mean	Standard Deviation	Minimum	25th Percentile	Median (50%)	75th Percentile	Maximum
AGE 9/GRADE 4 SPRING MAIN SAMPLES								
Booklet 28	1454.	2295.5	1035.8	1124.9	1777.5	1839.8	3487.2	5077.7
Booklets 1-10	5903.	369.2	198.9	118.5	240.0	283.2	461.6	1840.2
Booklets 11-17	4217.	869.6	488.2	265.4	548.9	681.7	1115.3	2555.3
Booklets 18-24	4098.	374.5	202.3	118.5	240.1	289.9	471.0	1840.2
Excluded students	1050.	220.8	84.3	119.3	178.9	182.5	221.2	973.4
AGE 13/GRADE 8 SPRING MAIN SAMPLES								
Booklet 25	1404.	2085.5	1129.7	726.0	1267.3	1638.2	2918.7	13880.3
Booklets 1-7	3993.	377.8	236.3	113.7	223.4	305.1	503.0	2520.0
Booklets 15-21	3945.	379.1	225.2	113.7	227.6	305.1	503.0	2120.2
Booklets 8-14	4010.	740.4	433.8	227.6	444.2	551.3	972.9	5090.7
Excluded students	821.	234.2	132.2	80.7	159.8	182.7	299.3	1508.4
AGE 17/GRADE 12 SPRING MAIN SAMPLES								
Booklet 25	1490.	1543.9	887.7	427.7	843.6	1299.7	2021.6	4211.3
Booklets 1-7	3984.	287.7	178.5	19.6	153.7	247.5	366.2	1426.3
Booklets 15-21	4037.	289.8	178.9	19.6	163.4	247.0	377.1	1426.3
Booklets 8-14	4186.	574.2	361.2	30.6	307.6	477.5	813.5	2208.3
Excluded students	639.	174.0	89.3	32.1	105.9	145.2	224.6	552.1

Table C-3

Distribution of Student Base Weights, Bridge Samples

Sample	Number of Cases	Mean	Standard Deviation	Minimum	25th Percentile	Median (50%)	75th Percentile	Maximum
AGE 9 BRIDGE SAMPLES								
Booklets 51-56	5926	667.0	245.4	213.8	576.6	666.3	757.8	2632.9
Booklets 91-93	6235	454.0	178.4	151.7	352.0	385.5	547.7	1771.9
Booklets 94-95	4134	670.3	248.7	245.9	576.6	664.4	756.5	2645.6
Excluded students	1116	259.7	194.0	75.9	174.1	189.8	200.1	1280.9
AGE 13 BRIDGE SAMPLES								
Booklets 51-56	6233	566.9	244.0	117.3	422.7	595.1	646.8	2179.2
Booklets 91-93	6649	360.6	150.0	115.5	294.8	321.6	367.6	1454.4
Booklets 94-95	4455	569.0	245.7	132.0	407.1	566.5	646.8	3427.4
Excluded students	1095	239.5	192.6	68.7	153.4	162.7	222.7	1755.6
AGE 17 BRIDGE SAMPLES								
Booklets 51-56	5614	419.0	185.5	9.0	328.1	366.0	455.8	1450.0
Booklets 61-66	8338	279.3	160.5	8.0	162.3	257.0	332.4	1180.2
Booklets 84-85	4411	424.2	185.5	9.0	331.4	371.2	460.3	1327.8
Booklets 94-95	4402	422.2	179.7	9.0	331.9	371.2	458.5	1327.8
Excluded students	1239	121.8	66.3	45.3	79.3	101.9	144.9	590.1

Table C-4

Distribution of School Nonresponse Adjustments, Winter Main Samples

Sample	Number of Cases	Mean	Standard Deviation	Minimum	25th Percentile	Median (50%)	75th Percentile	Maximum
AGE 9/GRADE 4 WINTER MAIN SAMPLES								
Booklet 28	1733	065	0.078	0.954	1.010	1.045	1.130	1.321
Booklets 1-10	6192	1.065	0.078	0.954	1.010	1.045	1.130	1.321
Booklets 11-17	4573	1.063	0.076	0.954	1.010	1.045	1.130	1.321
Booklets 18-24	4320	1.063	0.077	0.954	1.010	1.044	1.130	1.321
Excluded students	1282	1.064	0.065	0.954	1.010	1.061	1.061	1.321
AGE 13/GRADE 8 WINTER MAIN SAMPLES								
Booklet 25	1778	1.083	0.128	0.911	1.000	1.040	1.116	1.484
Booklets 1-7	4732	1.088	0.128	0.911	1.000	1.040	1.116	1.484
Booklets 15-21	4764	1.087	0.129	0.911	1.000	1.040	1.116	1.484
Booklets 8-14	4624	1.099	0.137	0.911	1.000	1.047	1.176	1.484
Excluded students	1129	1.083	0.122	0.911	1.005	1.054	1.116	1.484
AGE 17/GRADE 12 WINTER MAIN SAMPLES								
Booklet 25	1649	1.174	0.183	1.000	1.000	1.119	1.298	1.608
Booklets 1-7	4367	1.177	0.187	1.000	1.000	1.119	1.298	1.608
Booklets 15-21	4408	1.176	0.186	1.000	1.000	1.119	1.298	1.608
Booklets 8-14	4220	1.182	0.186	1.000	1.000	1.132	1.298	1.608
Excluded students	807	1.186	0.192	1.000	1.062	1.132	1.298	1.608

Table C-5

Distribution of School Nonresponse Adjustments, Spring Main Samples

Sample	Number of Cases	Mean	Standard Deviation	Minimum	25th Percentile	Median (50%)	75th Percentile	Maximum
AGE 9/GRADE 4 SPRING MAIN SAMPLES								
Booklet 28	1454	1.084	0.078	1.000	1.013	1.075	1.164	1.293
Booklets 1-10	5903	1.085	0.083	1.000	1.000	1.075	1.164	1.293
Booklets 11-17	4217	1.090	0.085	1.000	1.013	1.075	1.170	1.293
Booklets 18-24	4098	1.084	0.083	1.000	1.013	1.075	1.164	1.293
Excluded students	1050	1.106	0.103	1.000	1.000	1.078	1.170	1.293
AGE 13/GRADE 8 SPRING MAIN SAMPLES								
Booklet 25	1404	1.166	0.276	1.000	1.000	1.025	1.303	2.019
Booklets 1-7	3993	1.175	0.279	1.000	1.000	1.025	1.303	2.019
Booklets 15-21	3945	1.179	0.282	1.000	1.000	1.025	1.303	2.019
Booklets 8-14	4010	1.176	0.283	1.000	1.000	1.025	1.303	2.019
Excluded students	821	1.192	0.277	1.000	1.000	1.071	1.303	2.019
AGE 17/GRADE 12 SPRING MAIN SAMPLES								
Booklet 25	1490	1.234	0.232	1.000	1.048	1.195	1.376	2.144
Booklets 1-7	3984	1.225	0.225	1.000	1.048	1.195	1.376	2.144
Booklets 15-21	4037	1.225	0.230	1.000	1.034	1.187	1.376	2.144
Booklets 8-14	4186	1.220	0.221	1.000	1.034	1.187	1.376	2.144
Excluded students	639	1.206	0.198	1.000	1.048	1.195	1.373	2.144

Table C-6

Distribution of School Nonresponse Adjustments, Bridge Samples

Sample	Number of Cases	Mean	Standard Deviation	Minimum	25th Percentile	Median (50%)	75th Percentile	Maximum
AGE 9 BRIDGE SAMPLES								
Booklets 51-56	5926	1.102	0.081	1.000	1.023	1.088	1.171	1.263
Booklets 91-93	6235	1.098	0.075	1.000	1.024	1.088	1.168	1.238
Booklets 94-95	4134	1.097	0.075	1.000	1.024	1.088	1.168	1.238
Excluded students	1116	1.080	0.069	1.000	1.022	1.044	1.132	1.263
AGE 13 BRIDGE SAMPLES								
Booklets 51-56	6233	1.116	0.093	1.000	1.000	1.158	1.186	1.338
Booklets 91-93	6649	1.114	0.092	1.000	1.000	1.155	1.181	1.339
Booklets 94-95	4455	1.110	0.091	1.000	1.000	1.088	1.181	1.339
Excluded students	1095	1.114	0.088	1.000	1.000	1.158	1.186	1.339
AGE 17 BRIDGE SAMPLES								
Booklets 51-56	5614	1.159	0.105	1.000	1.099	1.149	1.178	1.427
Booklets 61-66	8338	1.162	0.106	1.000	1.099	1.149	1.178	1.427
Booklets 84-85	4411	1.162	0.104	1.000	1.099	1.149	1.178	1.427
Booklets 94-95	4402	1.162	0.105	1.000	1.099	1.149	1.178	1.427
Excluded students	1239	1.158	0.108	1.000	1.094	1.147	1.178	1.427

Table C-7

Distribution of Session Nonresponse Adjustments, Winter Main Samples

Sample	Number of Cases	Mean	Standard Deviation	Minimum	25th Percentile	Median (50%)	75th Percentile	Maximum
AGE 9/GRADE 4 WINTER MAIN SAMPLES								
Booklet 26	1733	1.029	0.044	1.000	1.000	1.000	1.076	1.131
Booklets 1-10	6192	1.029	0.062	0.852	1.000	1.029	1.058	1.228
Booklets 11-17	4573	1.037	0.048	1.000	1.000	1.026	1.055	1.239
Booklets 18-24	4320	1.030	0.067	0.852	1.000	1.029	1.058	1.228
Excluded students	1282	1.036	0.040	1.000	1.000	1.046	1.055	1.230
AGE 13/GRADE 8 WINTER MAIN SAMPLES								
Booklet 25	1778	1.005	0.030	1.000	1.000	1.000	1.000	1.174
Booklets 1-7	4732	1.007	0.024	1.000	1.000	1.000	1.000	1.102
Booklets 15-21	4764	1.007	0.024	1.000	1.000	1.000	1.000	1.102
Booklets 8-14	4624	1.015	0.025	1.000	1.000	1.000	1.017	1.089
Excluded students	1129	1.006	0.021	1.000	1.000	1.000	1.000	1.099
AGE 17/GRADE 12 WINTER MAIN SAMPLES								
Booklet 25	1649	1.026	0.050	1.000	1.000	1.000	1.051	1.175
Booklets 1-7	4367	1.008	0.019	1.000	1.000	1.000	1.006	1.073
Booklets 15-21	4408	1.008	0.019	1.000	1.000	1.000	1.006	1.073
Booklets 8-14	4220	1.024	0.030	1.000	1.000	1.000	1.051	1.080
Excluded students	807	1.005	0.015	1.000	1.000	1.000	1.000	1.070

571

572

Table C-8

Distribution of Session Nonresponse Adjustments, Spring Main Samples

Sample	Number of Cases	Mean	Standard Deviation	Minimum	25th Percentile	Median (50%)	75th Percentile	Maximum
AGE 9/GRADE 4 SPRING MAIN SAMPLES								
Booklet 28	1454	1.056	0.146	1.000	1.000	1.000	1.000	1.522
Booklets 1-10	5903	1.030	0.039	1.000	1.000	1.000	1.051	1.122
Booklets 11-17	4217	1.019	0.036	1.000	1.000	1.000	1.030	1.127
Booklets 18-24	4098	1.029	0.039	1.000	1.000	1.000	1.051	1.122
Excluded students	1050	1.032	0.037	1.000	1.000	1.019	1.047	1.123
AGE 13/GRADE 8 SPRING MAIN SAMPLES								
Booklet 25	1404	1.028	0.060	1.000	1.000	1.000	1.000	1.184
Booklets 1-7	3993	1.032	0.040	1.000	1.000	1.008	1.063	1.107
Booklets 15-21	3945	1.032	0.040	1.000	1.000	1.008	1.063	1.107
Booklets 8-14	4010	1.035	0.047	1.000	1.000	1.000	1.082	1.146
Excluded students	821	1.039	0.042	1.000	1.000	1.007	1.077	1.106
AGE 17/GRADE 12 SPRING MAIN SAMPLES								
Booklet 25	1490	1.000	0.000	1.000	1.000	1.000	1.000	1.000
Booklets 1-7	3984	1.001	0.006	1.000	1.000	1.000	1.000	1.041
Booklets 15-21	4037	1.001	0.006	1.000	1.000	1.000	1.000	1.041
Booklets 8-14	4186	1.003	0.009	1.000	1.000	1.000	1.000	1.041
Excluded students	639	1.000	0.003	1.000	1.000	1.000	1.000	1.041

Table C-9

Distribution of Session Nonresponse Adjustments, Bridge Samples

Sample	Number of Cases	Mean	Standard Deviation	Minimum	25th Percentile	Median (50%)	75th Percentile	Maximum
AGE 9 BRIDGE SAMPLES								
Booklets 51-56	5926	1.008	0.016	1.000	1.000	1.000	1.000	1.064
Booklets 91-93	6235	1.007	0.017	1.000	1.000	1.000	1.000	1.071
Booklets 94-95	4134	1.018	0.028	1.000	1.000	1.000	1.028	1.104
Excluded students	1116	1.007	0.012	1.000	1.000	1.000	1.016	1.075
AGE 13 BRIDGE SAMPLES								
Booklets 51-56	6233	1.019	0.031	1.000	1.000	1.000	1.026	1.100
Booklets 91-93	6649	1.025	0.027	1.000	1.000	1.015	1.051	1.069
Booklets 94-95	4455	1.023	0.048	1.000	1.000	1.000	1.010	1.145
Excluded students	1095	1.011	0.018	1.000	1.000	1.000	1.020	1.087
AGE 17 BRIDGE SAMPLES								
Booklets 51-56	5614	1.042	0.060	1.000	1.000	1.000	1.050	1.230
Booklets 61-66	8338	1.046	0.063	0.998	1.000	1.043	1.076	1.253
Booklets 84-85	4411	1.049	0.064	1.000	1.000	1.046	1.048	1.236
Booklets 94-95	4402	1.051	0.062	1.000	1.000	1.047	1.090	1.231
Excluded students	1239	1.036	0.051	1.000	1.000	1.009	1.050	1.212

Table C-10

Distribution of Age-only- and Grade-only-eligible Student Nonresponse Adjustments, Winter Main Samples

Sample	Number of Cases	Mean	Standard Deviation	Minimum	25th Percentile	Median (50%)	75th Percentile	Maximum
AGE 9/GRADE 4 WINTER MAIN SAMPLES								
Booklet 28	1733	1.021	0.078	1.000	1.000	1.000	1.000	1.506
Booklets 1-10	6192	1.010	0.051	1.000	1.000	1.000	1.000	1.480
Booklets 11-17	4573	1.009	0.049	1.000	1.000	1.000	1.000	1.496
Booklets 18-24	4320	1.010	0.050	1.000	1.000	1.000	1.000	1.480
Excluded students	1282	1.014	0.054	1.000	1.000	1.000	1.000	1.480
AGE 13/GRADE 8 WINTER MAIN SAMPLES								
Booklet 25	1778	1.006	0.031	1.000	1.000	1.000	1.000	1.174
Booklets 1-7	4732	1.007	0.030	1.000	1.000	1.000	1.000	1.199
Booklets 15-21	4764	1.007	0.031	1.000	1.000	1.000	1.000	1.199
Booklets 8-14	4624	1.005	0.028	1.000	1.000	1.000	1.000	1.195
Excluded students	1129	1.012	0.039	1.000	1.000	1.000	1.000	1.194
AGE 17/GRADE 12 WINTER MAIN SAMPLES								
Booklet 25	1649	1.009	0.036	1.000	1.000	1.000	1.000	1.181
Booklets 1-7	4367	1.004	0.026	1.000	1.000	1.000	1.000	1.187
Booklets 15-21	4408	1.003	0.024	1.000	1.000	1.000	1.000	1.187
Booklets 8-14	4220	1.002	0.022	1.000	1.000	1.000	1.000	1.205
Excluded students	807	1.004	0.018	1.000	1.000	1.000	1.000	1.181

Table C-11

Distribution of Age-only- and Grade-only-eligible Student Nonresponse Adjustments, Spring Main Samples

Sample	Number of Cases	Mean	Standard Deviation	Minimum	25th Percentile	Median (50%)	75th Percentile	Maximum
AGE 9/GRADE 4 SPRING MAIN SAMPLES								
Booklet 28	1454	1.017	0.065	1.000	1.000	1.000	1.000	1.443
Booklets 1-10	5903	1.010	0.034	1.000	1.000	1.000	1.000	1.279
Booklets 11-17	4217	1.008	0.031	1.000	1.000	1.000	1.000	1.269
Booklets 18-24	4098	1.010	0.034	1.000	1.000	1.000	1.000	1.279
Excluded students	1050	1.016	0.040	1.000	1.000	1.000	1.016	1.272
AGE 13/GRADE 8 SPRING MAIN SAMPLES								
Booklet 25	1404	1.011	0.047	1.000	1.000	1.000	1.000	1.276
Booklets 1-7	3993	1.008	0.032	1.000	1.000	1.000	1.000	1.283
Booklets 15-21	3945	1.008	0.034	1.000	1.000	1.000	1.000	1.283
Booklets 8-14	4010	1.009	0.035	1.000	1.000	1.000	1.000	1.210
Excluded students	821	1.013	0.040	1.000	1.000	1.000	1.000	1.219
AGE 17/GRADE 12 SPRING MAIN SAMPLES								
Booklet 25	1490	1.007	0.039	1.000	1.000	1.000	1.000	1.266
Booklets 1-7	3984	1.005	0.032	1.000	1.000	1.000	1.000	1.300
Booklets 15-21	4037	1.004	0.030	1.000	1.000	1.000	1.000	1.300
Booklets 8-14	4186	1.004	0.030	1.000	1.000	1.000	1.000	1.300
Excluded students	639	1.013	0.051	1.000	1.000	1.000	1.000	1.300

Table C-12

Distribution of Age-only- and Grade-only-eligible Student Nonresponse Adjustments, Bridge Samples

Sample	Number of Cases	Mean	Standard Deviation	Minimum	25th Percentile	Median (50%)	75th Percentile	Maximum
AGE 9 BRIDGE SAMPLES								
Booklets 51-56	5926	1.002	0.010	1.000	1.000	1.000	1.000	1.074
Booklets 91-93	6235	1.003	0.011	1.000	1.000	1.000	1.000	1.058
Booklets 94-95	4134	1.004	0.014	1.000	1.000	1.000	1.000	1.072
Excluded students	1116	1.004	0.012	1.000	1.000	1.000	1.000	1.060
AGE 13 BRIDGE SAMPLES								
Booklets 51-56	6233	1.007	0.027	1.000	1.000	1.000	1.000	1.209
Booklets 91-93	6649	1.006	0.022	[illegible]	1.000	1.000	1.000	1.121
Booklets 94-95	4455	1.005	0.017	1.000	1.000	1.000	1.000	1.092
Excluded students	1095	1.012	0.027	1.000	1.000	1.000	1.009	1.104
AGE 17 BRIDGE SAMPLES								
Booklets 51-56	5614	1.000	0.005	1.000	1.000	1.000	1.000	1.042
Booklets 61-66	8338	1.000	0.004	1.000	1.000	1.000	1.000	1.036
Booklets 84-85	4411	1.001	0.006	1.000	1.000	1.000	1.000	1.046
Booklets 94-95	4402	1.001	0.005	1.000	1.000	1.000	1.000	1.042
Excluded students	1239	1.001	0.006	1.000	1.000	1.000	1.000	1.039

Table C-13

Distribution of Student Nonresponse Adjustments, Winter Main Samples

Sample	Number of Cases	Mean	Standard Deviation	Minimum	25th Percentile	Median (50%)	75th Percentile	Maximum
AGE 9/GRADE 4 WINTER MAIN SAMPLES								
Booklet 28	1733	1.082	0.071	1.000	1.040	1.0625	1.105	1.312
Booklets 1-10	6192	1.079	0.059	1.000	1.039	1.075	1.100	1.419
Booklets 11-17	4573	1.078	0.062	1.000	1.031	1.065	1.100	1.363
Booklets 18-24	4320	1.080	0.059	1.000	1.041	1.074	1.099	1.419
Excluded students	1282	1.063	0.189	1.000	1.000	1.000	1.048	2.040
AGE 13/GRADE 8 WINTER MAIN SAMPLES								
Booklet 25	1778	1.132	0.110	1.000	1.044	1.105	1.182	1.440
Booklets 1-7	4732	1.128	0.098	1.000	1.054	1.103	1.182	1.518
Booklets 15-21	4764	1.129	0.099	1.000	1.054	1.108	1.186	1.518
Booklets 8-14	4624	1.136	0.113	1.000	1.067	1.108	1.171	1.714
Excluded students	1129	1.119	0.306	1.000	1.000	1.000	1.016	2.257
AGE 17/GRADE 12 WINTER MAIN SAMPLES								
Booklet 25	1649	1.218	0.175	1.000	1.091	1.186	1.292	1.834
Booklets 1-7	4367	1.246	0.185	1.000	1.125	1.192	1.310	2.358
Booklets 15-21	4408	1.246	0.189	1.000	1.120	1.191	1.309	2.358
Booklets 8-14	4220	1.257	0.187	1.021	1.141	1.233	1.310	2.086
Excluded students	807	1.062	0.161	1.000	1.000	1.000	1.000	1.739

Table C-14

Distribution of Student Nonresponse Adjustments, Spring Main Samples

Sample	Number of Cases	Mean	Standard Deviation	Minimum	25th Percentile	Median (50%)	75th Percentile	Maximum
AGE 9/GRADE 4 SPRING MAIN SAMPLES								
Booklet 28	1454	1.075	0.060	1.000	1.048	1.073	1.095	1.255
Booklets 1-10	5903	1.074	0.050	1.000	1.039	1.068	1.101	1.429
Booklets 11-17	4217	1.075	0.063	1.000	1.044	1.067	1.106	1.607
Booklets 18-24	4098	1.074	0.051	1.000	1.040	1.068	1.100	1.429
Excluded students	1050	1.043	0.094	1.000	1.000	1.000	1.072	1.386
AGE 13/GRADE 8 SPRING MAIN SAMPLES								
Booklet 25	1404	1.143	0.095	1.000	1.067	1.113	1.218	1.400
Booklets 1-7	3993	1.118	0.089	1.000	1.059	1.090	1.155	1.630
Booklets 15-21	3945	1.118	0.090	1.000	1.060	1.094	1.155	1.630
Booklets 8-14	4010	1.112	0.076	1.000	1.059	1.095	1.164	1.348
Excluded students	821	1.075	0.150	1.000	1.000	1.000	1.027	1.621
AGE 17/GRADE 12 SPRING MAIN SAMPLES								
Booklet 25	1490	1.224	0.141	1.000	1.143	1.200	1.260	1.845
Booklets 1-7	3984	1.238	0.155	1.000	1.150	1.200	1.278	2.030
Booklets 15-21	4037	1.240	0.153	1.000	1.153	1.200	1.278	2.030
Booklets 8-14	4186	1.244	0.155	1.000	1.146	1.229	1.295	2.073
Excluded students	639	1.043	0.125	1.000	1.000	1.000	1.005	1.475

Table C-15

Distribution of Student Nonresponse Adjustments, Bridge Samples

Sample	Number of Cases	Mean	Standard Deviation	Minimum	25th Percentile	Median (50%)	75th Percentile	Maximum
AGE 9 BRIDGE SAMPLES								
Booklets 51-56	5926	1.082	0.054	1.000	1.047	1.079	1.107	1.289
Booklets 91-93	6235	1.081	0.051	1.000	1.045	1.079	1.102	1.266
Booklets 94-95	4134	1.085	0.060	1.000	1.042	1.077	1.117	1.357
Excluded students	1116	1.033	0.100	1.000	1.000	1.000	1.000	1.425
AGE 13 BRIDGE SAMPLES								
Booklets 51-56	6233	1.102	0.081	1.000	1.045	1.074	1.137	1.500
Booklets 91-93	6649	1.113	0.108	1.000	1.047	1.080	1.162	1.891
Booklets 94-95	4455	1.110	0.101	1.000	1.044	1.08	1.155	1.888
Excluded students	1095	1.011	0.045	1.000	1.000	1.000	1.000	1.345
AGE 17 BRIDGE SAMPLES								
Booklets 51-56	5614	1.246	0.154	1.000	1.143	1.227	1.310	2.117
Booklets 61-66	8338	1.229	0.173	1.000	1.124	1.189	1.283	2.427
Booklets 84-85	4411	1.220	0.163	1.000	1.130	1.187	1.264	2.359
Booklets 94-95	4402	1.207	0.134	1.000	1.125	1.184	1.255	2.229
Excluded students	1239	1.050	0.103	1.000	1.000	1.000	1.013	1.294

Table C-16

Distribution of Trimming Factors, Winter Main Samples

Sample	Number of Cases	Mean	Standard Deviation	Minimum	25th Percentile	Median (50%)	75th Percentile	Maximum
AGE 9/GRADE 4 WINTER MAIN SAMPLES								
Booklet 28	1733	0.999	0.007	0.951	1.000	1.000	1.000	1.000
Booklets 1-10	6192	0.997	0.023	0.819	1.000	1.000	1.000	1.000
Booklets 11-17	4573	0.991	0.070	0.357	1.000	1.000	1.000	1.000
Booklets 18-24	4320	0.996	0.024	0.819	1.000	1.000	1.000	1.000
Excluded students	1282	0.967	0.115	0.538	1.000	1.000	1.000	1.000
AGE 13/GRADE 8 WINTER MAIN SAMPLES								
Booklet 25	1778	0.992	0.042	0.621	1.000	1.000	1.000	1.000
Booklets 1-7	4732	0.999	0.004	0.958	1.000	1.000	1.000	1.000
Booklets 15-21	4764	0.999	0.004	0.958	1.000	1.000	1.000	1.000
Booklets 8-14	4624	1.000	0.000	1.000	1.000	1.000	1.000	1.000
Excluded students	1129	0.997	0.018	0.879	1.000	1.000	1.000	1.000
AGE 17/GRADE 12 WINTER MAIN SAMPLES								
Booklet 25	1649	1.000	0.000	1.000	1.000	1.000	1.000	1.000
Booklets 1-7	4367	0.999	0.004	0.965	1.000	1.000	1.000	1.000
Booklets 15-21	4408	0.999	0.004	0.965	1.000	1.000	1.000	1.000
Booklets 8-14	4220	0.999	0.008	0.917	1.000	1.000	1.000	1.000
Excluded students	807	0.995	0.030	0.811	1.000	1.000	1.000	1.000

Table C-17

Distribution of Trimming Factors, Spring Main Samples

Sample	Number of Cases	Mean	Standard Deviation	Minimum	25th Percentile	Median (50%)	75th Percentile	Maximum
AGE 9/GRADE 4 SPRING MAIN SAMPLES								
Booklet 28	1454	1.000	0.000	1.000	1.000	1.000	1.000	1.000
Booklets 1-10	5903	0.989	0.051	0.680	1.000	1.000	1.000	1.000
Booklets 11-17	4217	0.996	0.023	0.816	1.000	1.000	1.000	1.000
Booklets 18-24	4098	0.989	0.052	0.680	1.000	1.000	1.000	1.000
Excluded students	1050	0.961	0.124	0.501	1.000	1.000	1.000	1.000
AGE 13/GRADE 8 SPRING MAIN SAMPLES								
Booklet 25	1404	1.000	0.000	1.000	1.000	1.000	1.000	1.000
Booklets 1-7	3993	0.995	0.026	0.804	1.000	1.000	1.000	1.000
Booklets 15-21	3945	0.995	0.028	0.804	1.000	1.000	1.000	1.000
Booklets 8-14	4010	0.998	0.016	0.865	1.000	1.000	1.000	1.000
Excluded students	821	0.973	0.095	0.606	1.000	1.000	1.000	1.000
AGE 17/GRADE 12 SPRING MAIN SAMPLES								
Booklet 25	1490	0.996	0.025	0.833	1.000	1.000	1.000	1.000
Booklets 1-7	3984	0.999	0.007	0.923	1.000	1.000	1.000	1.000
Booklets 15-21	4037	0.999	0.008	0.923	1.000	1.000	1.000	1.000
Booklets 8-14	4186	0.998	0.022	0.795	1.000	1.000	1.000	1.000
Excluded students	639	0.999	0.003	0.979	1.000	1.000	1.000	1.000

Table C-18

Distribution of Trimming Factors, Bridge Samples

Sample	Number of Cases	Mean	Standard Deviation	Minimum	25th Percentile	Median (50%)	75th Percentile	Maximum
AGE 9 BRIDGE SAMPLES								
Booklets 51-56	5926	0.996	0.037	0.559	1.000	1.000	1.000	1.000
Booklets 91-93	6235	0.997	0.035	0.622	1.000	1.000	1.000	1.000
Booklets 94-95	4134	0.997	0.035	0.607	1.000	1.000	1.000	1.000
Excluded students	1116	0.957	0.120	0.572	1.000	1.000	1.000	1.000
AGE 13 BRIDGE SAMPLES								
Booklets 51-56	6233	0.998	0.017	0.812	1.000	1.000	1.000	1.000
Booklets 91-93	6649	1.000	0.000	1.000	1.000	1.000	1.000	1.000
Booklets 94-95	4455	0.999	0.008	0.910	1.000	1.000	1.000	1.000
Excluded students	1095	0.956	0.133	0.485	1.000	1.000	1.000	1.000
AGE 17 BRIDGE SAMPLES								
Booklets 51-56	5614	0.996	0.040	0.521	1.000	1.000	1.000	1.000
Booklets 61-66	8338	0.998	0.028	0.561	1.000	1.000	1.000	1.000
Booklets 84-85	4411	0.994	0.042	0.576	1.000	1.000	1.000	1.000
Booklets 94-95	4402	0.996	0.036	0.659	1.000	1.000	1.000	1.000
Excluded students	1239	0.967	0.110	0.470	1.000	1.000	1.000	1.000

Table C-19

Distribution of Poststratification Factors, Winter Main Samples

Sample	Number of Cases	Mean	Standard Deviation	Minimum	25th Percentile	Median (50%)	75th Percentile	Maximum
AGE 9/GRADE 4 WINTER MAIN SAMPLES								
Booklet 28	1733	0.956	0.311	0.516	0.766	0.908	1.170	1.611
Booklets 1-10	6192	0.999	0.288	0.553	0.771	1.078	1.232	1.561
Booklets 11-17	4573	0.969	0.257	0.536	0.713	1.039	1.223	1.382
Booklets 18-24	4320	0.990	0.289	0.553	0.718	1.078	1.232	1.561
Excluded students	1282	0.943	0.324	0.559	0.566	0.910	1.206	1.458
AGE 13/GRADE 8 WINTER MAIN SAMPLES								
Booklet 25	1778	0.978	0.208	0.476	0.812	1.011	1.075	1.544
Booklets 1-7	4732	0.970	0.184	0.616	0.842	0.944	1.083	1.269
Booklets 15-21	4764	0.970	0.186	0.616	0.836	0.944	1.152	1.269
Booklets 8-14	4624	1.006	0.219	0.566	0.850	1.021	1.133	1.553
Excluded students	1129	0.989	0.215	0.602	0.797	0.940	1.202	1.342
AGE 17/GRADE 12 WINTER MAIN SAMPLES								
Booklet 25	1649	0.985	0.229	0.653	0.797	1.066	1.123	1.369
Booklets 1-7	4367	1.028	0.220	0.708	0.785	1.059	1.213	1.412
Booklets 15-21	4408	1.027	0.221	0.708	0.785	1.059	1.213	1.412
Booklets 8-14	4220	1.005	0.186	0.665	0.834	1.058	1.187	1.260
Excluded students	807	0.988	0.209	0.714	0.804	0.915	1.210	1.335

Table C-20

Distribution of Poststratification Factors, Spring Main Samples

Sample	Number of Cases	Mean	Standard Deviation	Minimum	25th Percentile	Median (50%)	75th Percentile	Maximum
AGE 9/GRADE 4 SPRING MAIN SAMPLES								
Booklet 28	1454	1.167	0.360	0.381	1.012	1.075	1.447	2.789
Booklets 1-10	5903	1.100	0.248	0.569	1.032	1.161	1.273	1.505
Booklets 11-17	4217	1.099	0.278	0.558	0.908	1.235	1.279	1.618
Booklets 18-24	4098	1.113	0.244	0.569	1.032	1.161	1.273	1.505
Excluded students	1050	1.085	0.291	0.552	0.685	1.221	1.328	1.454
AGE 13/GRADE 8 SPRING MAIN SAMPLES								
Booklet 25	1404	1.144	0.408	0.706	0.794	1.061	1.285	2.203
Booklets 1-7	3993	1.141	0.303	0.690	0.895	1.164	1.397	1.931
Booklets 15-21	3945	1.130	0.297	0.690	0.882	1.045	1.397	0.931
Booklets 8-14	4010	1.140	0.297	0.727	0.896	1.152	1.347	1.754
Excluded students	821	1.173	0.312	0.702	0.878	1.122	1.426	1.822
AGE 17/GRADE 12 SPRING MAIN SAMPLES								
Booklet 25	1490	1.114	0.268	0.791	0.907	1.042	1.360	1.674
Booklets 1-7	3984	1.118	0.186	0.878	0.998	1.138	1.221	1.532
Booklets 15-21	4037	1.113	0.186	0.878	0.998	1.134	1.221	1.532
Booklets 8-14	4186	1.074	0.163	0.841	0.942	1.088	1.206	1.452
Excluded students	639	1.150	0.167	0.862	1.010	1.153	1.237	1.451

Table C-21

Distribution of Poststratification Factors, Bridge Samples

Sample	Number of Cases	Mean	Standard Deviation	Minimum	25th Percentile	Median (50%)	75th Percentile	Maximum
AGE 9 BRIDGE SAMPLES								
Booklets 51-56	5926	1.026	0.234	0.595	0.913	1.012	1.187	1.501
Booklets 91-93	6235	1.077	0.227	0.615	1.149	1.164	1.223	1.276
Booklets 94-95	4134	1.072	0.255	0.599	0.795	1.209	1.232	1.286
Excluded students	1116	1.047	0.299	0.625	0.646	1.129	1.306	1.427
AGE 13 BRIDGE SAMPLES								
Booklets 51-56	6233	0.992	0.188	0.651	0.905	0.972	1.077	1.537
Booklets 91-93	6649	1.021	0.165	0.853	0.856	1.011	1.097	1.394
Booklets 94-95	4455	0.974	0.126	0.685	0.874	0.971	1.104	1.128
Excluded students	1095	1.006	0.185	0.683	0.914	0.972	1.081	1.537
AGE 17 BRIDGE SAMPLES								
Booklets 51-56	5614	1.171	0.228	0.876	0.968	1.136	1.360	1.696
Booklets 61-66	8338	1.193	0.225	0.903	1.031	1.191	1.263	1.710
Booklets 84-85	4411	1.102	0.102	0.959	1.044	1.067	1.197	1.234
Booklets 94-95	4402	1.112	0.112	0.988	1.029	1.051	1.265	1.307
Excluded students	1239	1.289	0.211	0.918	1.121	1.303	1.420	1.705

Table C-22

Distribution of Aggregate Adjustments to Base Weights, Winter Main Samples

Sample	Number of Cases	Mean	Standard Deviation	Minimum	25th Percentile	Median (50%)	75th Percentile	Maximum
AGE 9/GRADE 4 WINTER MAIN SAMPLES								
Booklet 28	1733.	1.155	0.412	0.493	0.846	1.103	1.461	2.780
Booklets 1-10	6192.	1.193	0.389	0.518	0.839	1.242	1.454	2.909
Booklets 11-17	4573.	1.153	0.362	0.236	0.828	1.177	1.406	2.806
Booklets 18-24	4320.	1.181	0.392	0.518	0.823	1.232	1.449	2.909
Excluded students	1282.	1.104	0.515	0.350	0.659	1.077	1.429	3.285
AGE 13/GRADE 8 WINTER MAIN SAMPLES								
Booklet 25	1778.	1.228	0.331	0.428	0.990	1.194	1.409	378.0
Booklets 1-7	4732.	1.208	0.299	0.616	0.995	1.180	1.371	273.0
Booklets 15-21	4764.	1.208	0.302	0.616	0.995	1.180	1.371	2.730
Booklets 8-14	4624.	1.277	0.326	0.515	1.057	1.225	1.461	2.671
Excluded students	1129.	1.216	0.443	0.606	0.912	1.145	1.339	3.259
AGE 17/GRADE 12 WINTER MAIN SAMPLES								
Booklet 25	1649.	1.476	0.540	0.653	1.090	1.446	1.719	4.242
Booklets 1-7	4367.	1.532	0.500	0.708	1.171	1.444	1.765	2.596
Booklets 15-21	4408.	1.527	0.500	0.751	1.165	1.440	1.778	3.836
Booklets 8-14	4220.	1.542	0.493	0.751	1.208	1.463	1.759	3.911
Excluded students	807.	1.250	0.398	0.753	0.923	1.187	1.456	3.131

Table C-23

Distribution of Aggregate Adjustments to Base Weights, Spring Main Samples

Sample	Number of Cases	Mean	Standard Deviation	Minimum	25th Percentile	Median (50%)	75th Percentile	Maximum
AGE 9/GRADE 4 SPRING MAIN SAMPLES								
Booklet 28	1454.	1.460	0.546	0.386	1.116	1.323	1.736	6.127
Booklets 1-10	5903.	1.317	0.332	0.498	1.080	1.325	1.564	2.312
Booklets 11-17	4217.	1.317	0.365	0.540	1.111	1.352	1.532	3.364
Booklets 18-24	4098.	1.330	0.332	0.505	1.098	1.331	1.568	2.312
Excluded students	1050.	1.265	0.418	0.339	0.887	1.316	1.550	2.537
AGE 13/GRADE 8 SPRING MAIN SAMPLES								
Booklet 25	1404.	1.600	0.744	0.737	1.040	1.399	2.037	4.814
Booklets 1-7	3993.	1.569	0.630	0.642	1.115	1.436	1.871	4.588
Booklets 15-21	3945.	1.562	0.641	0.642	1.098	1.422	1.862	4.588
Booklets 8-14	4010.	1.565	0.606	0.690	1.120	1.454	1.901	4.034
Excluded students	821.	1.544	0.599	0.544	1.064	1.453	1.896	4.114
AGE 17/GRADE 12 SPRING MAIN SAMPLES								
Booklet 25	1490.	1.714	0.772	0.844	1.321	1.569	1.887	6.621
Booklets 1-7	3984.	1.729	0.646	0.889	1.388	1.599	1.856	6.049
Booklets 15-21	4037.	1.723	0.655	0.898	1.386	1.599	1.849	6.049
Booklets 8-14	4186.	1.656	0.592	0.869	1.360	1.555	1.849	6.452
Excluded students	639.	1.456	0.332	0.862	1.247	1.398	1.622	3.111

Table C-24

Distribution of Aggregate Adjustments to Base Weights, Bridge Samples

Sample	Number of Cases	Mean	Standard Deviation	Minimum	25th Percentile	Median (50%)	75th Percentile	Maximum
AGE 9 BRIDGE SAMPLES								
Booklets 51-56	5926	1.232	0.311	0.441	1.017	1.239	1.445	2.021
Booklets 91-93	6235	1.286	0.295	0.449	1.258	1.362	1.462	1.945
Booklets 94-95	4134	1.300	0.337	0.426	1.047	1.42	1.521	1.992
Excluded students	1116	1.152	0.416	0.383	0.689	1.293	1.509	2.0322
AGE 13 BRIDGE SAMPLES								
Booklets 51-56	6233	1.253	0.297	0.612	1.093	1.216	1.388	2.666
Booklets 91-93	6649	1.312	0.308	0.866	1.133	1.259	1.453	3.670
Booklets 94-95	4455	1.236	0.238	0.685	1.076	1.229	1.380	2.253
Excluded students	1095	1.117	0.306	0.424	0.927	1.112	1.268	2.648
AGE 17 BRIDGE SAMPLES								
Booklets 51-56	5614	1.772	0.521	0.787	1.377	1.701	2.060	3.960
Booklets 61-66	8338	1.795	0.546	0.851	1.434	1.677	2.052	5.930
Booklets 84-85	4411	1.642	0.395	0.969	1.408	1.546	1.823	4.169
Booklets 94-95	4402	1.639	0.348	0.928	1.382	1.582	1.814	4.172
Excluded students	1239	1.567	0.398	0.811	1.303	1.519	1.726	2.988

Table C-25

Distribution of Final Student Weights, Winter Main Samples

Sample	Number of Cases	Mean	Standard Deviation	Minimum	25th Percentile	Median (50%)	75th Percentile	Maximum
AGE 9/GRADE 4 WINTER MAIN SAMPLES								
Booklet 28	1733	2766.1	1626.2	637.7	1490.4	2331.3	3541.2	10369.1
Booklets 1-10	6192	457.3	318.0	68.1	233.7	369.2	573.4	2878.3
Booklets 11-17	4573	1051.2	695.4	162.0	573.9	791.4	1360.5	4178.4
Booklets 18-24	4320	454.5	312.2	68.1	234.5	367.3	566.5	2878.3
Excluded students	1282	274.1	192.6	63.3	118.6	223.2	400.6	1594.7
AGE 13/GRADE 8 WINTER MAIN SAMPLES								
Booklet 25	1778	2478.4	1392.6	560.9	1552.3	2032.6	3162.9	11764.8
Booklets 1-7	4732	464.0	257.3	74.5	287.6	406.5	593.5	3260.8
Booklets 15-21	4764	466.7	258.9	80.1	279.6	406.6	595.5	2954.8
Booklets 8-14	4624	952.7	534.6	193.6	599.3	841.6	1213.1	8451.1
Excluded students	1129	275.2	188.9	74.9	168.3	222.2	327.9	2346.2
AGE 17/GRADE 12 WINTER MAIN SAMPLES								
Booklet 25	1649	2297.2	1383.8	542.3	1259.1	2094.4	2909.9	10284.9
Booklets 1-7	4367	442.9	252.5	65.8	255.2	395.1	600.4	1487.6
Booklets 15-21	4408	436.4	242.4	39.9	252.5	395.1	593.8	1486.6
Booklets 8-14	4220	896.2	539.6	160.6	498.2	799.5	1102.5	5457.0
Excluded students	807	230.1	123.2	57.5	142.1	195.8	290.6	764.4

Table C-26

Distribution of Final Student Weights, Spring Main Samples

Sample	Number of Cases	Mean	Standard Deviation	Minimum	25th Percentile	Median (50%)	75th Percentile	Maximum
AGE 9/GRADE 4 SPRING MAIN SAMPLES								
Booklet 28	1454	3308.5	1843.1	699.8	1975.6	2602.2	4373.3	12254.1
Booklets 1-10	5903	480.8	282.3	82.8	293.6	401.6	587.2	2632.3
Booklets 11-17	4217	1150.8	732.8	176.9	632.8	921.4	1463.1	4932.9
Booklets 18-24	4098	492.0	287.9	82.8	300.7	413.4	610.4	2770.9
Excluded students	1050	284.6	163.5	75.0	161.9	248.4	312.2	1307.5
AGE 13/GRADE 8 SPRING MAIN SAMPLES								
Booklet 25	1404	3159.3	2100.7	716.7	1839.0	2721.147	3907.9	37981.0
Booklets 1-7	3993	560.3	378.4	93.6	322.2	467.6	709.0	5891.8
Booklets 15-21	3945	561.0	356.6	87.4	324.6	467.9	715.7	4193.1
Booklets 8-14	4010	1108.5	727.1	252.7	637.7	914.2	1366.6	7833.0
Excluded students	821	337.0	192.4	81.9	232.9	292.1	403.6	2205.2
AGE 17/GRADE 12 SPRING MAIN SAMPLES								
Booklet 25	1490	2581.4	1668.7	467.4	1269.0	2126.6	3413.0	11964.6
Booklets 1-7	3984	481.3	309.1	20.5	254.5	411.8	640.6	2203.9
Booklets 15-21	4037	482.6	309.3	20.3	265.9	412.6	640.6	2203.9
Booklets 8-14	4186	917.8	571.6	27.9	473.4	792.3	1256.7	3776.4
Excluded students	639	248.5	128.3	56.9	148.1	197.5	333.1	779.7

Table C-27

Distribution of Final Student Weights, Bridge Samples

Sample	Number of Cases	Mean	Standard Deviation	Minimum	25th Percentile	Median (50%)	75th Percentile	Maximum
AGE 9 BRIDGE SAMPLES								
Booklets 51-56	5926	814.8	353.4	168.6	567.2	796.8	1010.0	3917.7
Booklets 91-93	6235	583.4	264.9	117.8	409.1	540.8	703.7	2652.2
Booklets 94-95	4134	870.8	390.0	168.0	545.8	907.8	1088.4	3908.0
Excluded students	1116	303.4	264.6	63.9	136.0	253.9	301.4	1632.4
AGE 13 BRIDGE SAMPLES								
Booklets 51-56	6233	704.3	341.0	118.9	498.3	675.1	844.9	3780.9
Booklets 91-93	6649	467.1	205.4	144.0	335.0	413.2	534.5	2525.5
Booklets 94-95	4455	699.5	304.6	111.4	488.6	703.8	845.8	3319.2
Excluded students	1095	271.9	250.2	59.7	141.8	179.8	255.6	1735.6
AGE 17 BRIDGE SAMPLES								
Booklets 51-56	5614	728.1	335.4	21.1	498.9	663.1	872.4	3007.5
Booklets 61-66	8338	489.8	287.6	25.0	289.4	431.4	604.9	2909.8
Booklets 84-85	4411	680.7	281.4	13.7	512.0	611.7	776.2	2794.8
Booklets 94-95	4402	681.6	283.6	15.56	500.1	614.6	781.8	2797.0
Excluded students	1239	191.5	117.2	47.56	116.8	150.4	232.3	890.3

APPENDIX D

Design Effects Statistics for 1990 NAEP Samples

Table D-1

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics
Across Cognitive Reading Cross-sectional Items
in the [Rdg-MainP] Samples

Grade 4 (Booklets 1-7)*

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.20	1.46	1.69	2.27	1.45	0.28	52
Male	1.16	1.40	1.63	2.08	1.41	0.31	40
Female	1.22	1.43	1.62	2.28	1.42	0.32	38
White	1.14	1.36	1.60	2.38	1 39	0.30	42
Black	1.01	1.17	1.41	2.30	1.21	0.28	37
Hispanic	1.10	1.22	1.44	2.47	1.29	0.28	40
Asian American	0.93	1.12	1.58	5.27	1.40	0.80	6
Other Race/Ethnicity	1.02	1.30	1.53	2.19	1.30	0.34	29
Other Metro	1.19	1.40	1.68	2.16	1.47	0.32	41
Disadvantaged Urban	1.04	1.27	1.52	2.33	1.30	0.37	25
Advantaged Urban	1.05	1.33	1.68	3.20	1.41	0.50	15
Par. Ed. < HS	1.03	1.22	1.62	2.95	1.31	0.41	20
Par. Ed. = HS	1.06	1.30	1.64	3.70	1.43	0.54	14
Par. Ed. > HS	1.19	1.35	1.55	2.44	1.40	0.32	39
Par. Ed. = Coll.	1.19	1.33	1.53	2.35	1.37	0.30	42
Par. Ed. = IDK	1.09	1.33	1.53	2.18	1.33	0.32	33
Public School	1.16	1.42	1.60	2.09	1.39	0.29	47
Nonpublic School	1.31	1.46	1.66	2.29	1.48	0.31	44

* Distributions are based on 67 items.

Table D-2

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics
Across Cognitive Reading Cross-sectional Items
in the [Rdg-MainP] Samples

Grade 8 (Booklets 1-7)*

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.30	1.47	1.73	2.65	1.53	0.35	37
Male	1.25	1.45	1.68	2.66	1.46	0.34	37
Female	1.19	1.35	1.56	2.12	1.37	0.27	49
White	1.20	1.41	1.64	2.83	1.43	0.34	35
Black	1.01	1.17	1.40	2.09	1.24	0.30	33
Hispanic	1.12	1.33	1.57	2.47	1.38	0.37	28
Asian American	1.01	1.28	1.66	4.39	1.44	0.64	10
Other Race/Ethnicity	0.92	1.10	1.49	4.51	1.31	0.67	8
Other Metro	1.27	1.46	1.64	2.55	1.49	0.36	34
Disadvantaged Urban	1.11	1.57	1.89	3.68	1.61	0.65	12
Advantaged Urban	0.78	1.18	1.81	6.56	1.48	1.01	4
Par. Ed. < HS	1.01	1.25	1.50	2.39	1.27	0.33	29
Par. Ed. = HS	0.99	1.16	1.45	2.30	1.23	0.34	25
Par. Ed. > HS	1.10	1.22	1.36	2.55	1.24	0.29	37
Par. Ed. = Coll.	1.12	1.34	1.56	2.25	1.36	0.31	37
Par. Ed. = IDK	1.16	1.29	1.51	2.95	1.34	0.32	34
Public School	1.17	1.39	1.65	2.25	1.42	0.34	35
Nonpublic School	1.31	1.54	1.84	2.68	1.56	0.42	27

* Distributions are based on 97 items.

Table D-3

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics
Across Cognitive Reading Cross-sectional Items
in the [Rdg-MainP] Samples

Grade 12 (Booklets 1-7)*

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.26	1.54	1.80	2.76	1.56	0.38	33
Male	1.20	1.47	1.69	2.72	1.48	0.37	32
Female	1.13	1.31	1.52	2.16	1.32	0.29	40
White	1.21	1.46	1.79	2.33	1.50	0.36	34
Black	1.03	1.22	1.52	2.58	1.27	0.34	28
Hispanic	1.14	1.41	1.64	2.86	1.43	0.38	27
Asian American	0.83	1.15	1.76	4.38	1.43	0.81	6
Other Race/Ethnicity	0.96	1.12	1.47	2.81	1.22	0.41	17
Other Metro	1.21	1.49	1.89	2.63	1.56	0.42	27
Disadvantaged Urban	1.26	1.73	2.32	7.45	1.90	1.07	6
Advantaged Urban	0.98	1.36	1.83	3.14	1.44	0.58	12
Par. Ed. < HS	0.98	1.15	1.43	2.11	1.21	0.29	34
Par. Ed. = HS	1.00	1.19	1.52	2.23	1.27	0.36	25
Par. Ed. > HS	1.12	1.36	1.55	2.56	1.37	0.33	34
Par. Ed. = Coll.	1.17	1.31	1.55	2.33	1.35	0.30	39
Par. Ed. = IDK	1.15	1.34	1.53	2.54	1.38	0.34	32
Public School	1.13	1.39	1.70	2.37	1.43	0.36	32
Nonpublic School	1.38	1.66	2.04	3.25	1.73	0.50	24

* Distributions are based on 112 items.

Table D-4

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics
Across Cognitive Reading Trend Items
in the [RW-Br84] Samples

Age 9 (Booklets 51-56)*

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.05	1.24	1.48	2.25	1.28	0.31	34
Male	1.02	1.14	1.35	1.92	1.20	0.30	31
Female	0.98	1.21	1.43	2.28	1.23	0.31	32
White	0.99	1.17	1.38	2.42	1.21	0.31	30
Black	1.06	1.31	1.55	2.92	1.35	0.42	21
Hispanic	0.92	1.07	1.27	2.61	1.10	0.31	25
Asian American	0.73	1.25	1.87	6.40	1.49	1.06	4
Other Race/Ethnicity	0.88	1.20	1.45	3.69	1.25	0.52	11
Other Metro	0.98	1.18	1.39	2.24	1.22	0.32	28
Disadvantaged Urban	1.15	1.54	1.98	5.82	1.66	0.78	9
Advantaged Urban	0.90	1.19	1.44	2.29	1.23	0.42	17
Par. Ed. < HS	0.99	1.14	1.36	2.07	1.19	0.29	33
Par. Ed. = HS	0.92	1.13	1.27	2.18	1.14	0.30	28
Par. Ed. > HS	0.88	1.07	1.28	2.12	1.11	0.31	26
Par. Ed. = Coll.	1.06	1.25	1.51	1.93	1.28	0.29	37
Par. Ed. = IDK	1.07	1.24	1.42	2.13	1.25	0.28	40
Public School	1.13	1.30	1.49	2.33	1.32	0.32	33
Nonpublic School	0.89	1.15	1.47	3.34	1.24	0.48	14

* Distributions are based on 105 items.

Table D-5

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics
Across Cognitive Reading Trend Items
in the [RW-Br84] Samples

Age 13 (Booklets 51-56)*

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.05	1.23	1.41	2.02	1.24	0.26	45
Male	1.06	1.20	1.42	1.99	1.25	0.30	34
Female	0.99	1.19	1.32	1.97	1.18	0.25	43
White	1.01	1.15	1.34	2.07	1.20	0.27	40
Black	1.04	1.33	1.57	2.55	1.35	0.38	25
Hispanic	1.03	1.28	1.60	2.43	1.34	0.42	20
Asian American	0.92	1.35	2.47	7.26	1.90	1.44	3
Other Race/Ethnicity	0.82	1.06	1.32	3.13	1.10	0.37	18
Other Metro	1.08	1.24	1.43	2.32	1.25	0.29	38
Disadvantaged Urban	1.00	1.30	1.66	3.13	1.40	0.61	10
Advantaged Urban	1.01	1.38	1.76	3.13	1.41	0.53	14
Par. Ed. < HS	1.02	1.19	1.39	2.08	1.23	0.30	34
Par. Ed. = HS	1.02	1.17	1.35	2.06	1.20	0.26	42
Par. Ed. > HS	0.91	1.10	1.40	2.82	1.19	0.37	21
Par. Ed. = Coll.	1.05	1.26	1.47	2.17	1.27	0.29	39
Par. Ed. = IDK	1.02	1.21	1.49	2.30	1.26	0.33	29
Public School	1.06	1.22	1.39	2.17	1.24	0.25	49
Nonpublic School	0.93	1.14	1.51	2.21	1.22	0.36	23

* Distributions are based on 107 items.

Table D-6

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics
Across Cognitive Reading Trend Items
in the [RW-Br84] Samples

Age 17 (Booklets 51-56)*

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.05	1.27	1.46	2.27	1.30	0.32	33
Male	1.01	1.21	1.50	2.37	1.28	0.33	30
Female	1.04	1.21	1.43	1.87	1.24	0.28	40
White	1.03	1.19	1.42	2.40	1.24	0.35	25
Black	1.02	1.16	1.47	2.55	1.25	0.35	26
Hispanic	0.88	1.14	1.44	2.57	1.18	0.44	15
Asian American	0.67	0.97	1.39	2.97	1.06	0.50	9
Other Race/Ethnicity	0.87	1.11	1.34	3.23	1.17	0.48	12
Other Metro	1.00	1.20	1.48	2.27	1.25	0.34	27
Disadvantaged Urban	0.86	1.20	1.66	3.25	1.34	0.65	8
Advantaged Urban	0.94	1.24	1.60	3.00	1.35	0.56	12
Par. Ed. < HS	0.94	1.20	1.34	4.46	1.21	0.47	13
Par. Ed. = HS	1.00	1.16	1.39	2.30	1.23	0.33	27
Par. Ed. > HS	1.01	1.22	1.45	2.27	1.25	0.33	28
Par. Ed. = Coll.	0.95	1.16	1.42	2.24	1.19	0.32	27
Par. Ed. = IDK	0.99	1.18	1.46	2.57	1.26	0.43	17
Public School	1.02	1.27	1.48	2.37	1.29	0.33	30
Nonpublic School	0.90	1.12	1.46	2.32	1.20	0.41	17

* Distributions are based on 95 items.

Table D-7

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics
Across Cognitive Reading Trend Items
in the [RMS-Br86] Samples

Age 9 (Booklets 91-93)*

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.62	1.97	2.20	2.67	1.91	0.42	39
Male	1.37	1.56	1.90	2.51	1.63	0.38	35
Female	1.32	1.55	1.77	2.43	1.57	0.30	53
White	1.37	1.60	1.75	2.06	1.54	0.29	55
Black	1.19	1.46	2.01	2.45	1.54	0.45	22
Hispanic	1.17	1.37	1.61	2.96	1.43	0.49	16
Asian American	1.00	1.43	3.04	6.51	2.16	1.57	4
Other Race/Ethnicity	0.87	0.96	1.12	1.71	1.01	0.24	35
Other Metro	1.44	1.72	1.91	2.41	1.70	0.30	56
Disadvantaged Urban	1.65	2.27	3.53	7.55	2.85	1.75	5
Advantaged Urban	0.91	1.35	1.81	4.31	1.53	0.79	7
Par. Ed. < HS	0.94	1.12	1.26	1.69	1.13	0.24	42
Par. Ed. = HS	1.03	1.40	1.62	2.13	1.36	0.33	33
Par. Ed. > HS	0.97	1.26	1.43	1.76	1.24	0.28	39
Par. Ed. = Coll.	1.48	1.68	1.92	2.24	1.64	0.33	47
Par. Ed. = IDK	1.24	1.42	1.69	2.23	1.46	0.33	38
Public School	1.48	1.83	2.19	2.56	1.83	0.40	41
Nonpublic School	1.05	1.47	1.99	6.40	1.85	1.32	4

* Distributions are based on 30 items.

Table D-8

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics
Across Cognitive Reading Trend Items
in the [RMS-Br86] Samples

Age 13 (Booklets 91-93)*

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.47	1.87	2.58	3.15	1.95	0.63	19
Male	1.30	1.54	1.95	2.88	1.62	0.50	21
Female	1.28	1.58	1.89	3.07	1.64	0.44	26
White	1.41	1.61	2.01	2.95	1.66	0.51	21
Black	0.92	1.19	1.61	2.89	1.32	0.52	12
Hispanic	1.15	1.36	1.60	2.83	1.40	0.44	20
Asian American	0.93	1.77	2.72	7.93	2.23	1.71	3
Other Race/Ethnicity	0.81	0.99	1.29	1.77	1.06	0.32	21
Other Metro	1.52	1.85	2.24	2.87	1.82	0.56	20
Disadvantaged Urban	1.08	1.57	1.89	9.41	1.82	1.49	3
Advantaged Urban	0.84	1.16	1.62	3.17	1.30	0.63	8
Par. Ed. < HS	0.94	1.11	1.50	2.45	1.23	0.44	15
Par. Ed. = HS	1.08	1.34	1.54	1.97	1.31	0.30	37
Par. Ed. > HS	1.14	1.34	1.54	2.21	1.37	0.31	39
Par. Ed. = Coll.	1.17	1.63	1.81	2.63	1.58	0.45	24
Par. Ed. = IDK	1.03	1.18	1.56	1.87	1.27	0.31	33
Public School	1.41	2.00	2.56	3.30	1.95	0.71	15
Nonpublic School	0.71	1.13	1.98	4.12	1.37	0.92	4

* Distributions are based on 34 items.

Table D-9

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics
Across Cognitive Reading Trend Items
in the [RMS-Br86] Samples

Age 17 (Booklets 61-66)*

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.26	1.42	1.68	2.70	1.47	0.34	37
Male	1.24	1.40	1.73	2.80	1.46	0.34	36
Female	1.23	1.38	1.76	2.67	1.50	0.43	24
White	1.14	1.32	1.56	2.60	1.37	0.36	29
Black	1.20	1.54	1.94	3.29	1.64	0.59	15
Hispanic	0.98	1.19	1.68	2.81	1.35	0.48	16
Asian American	0.78	1.00	1.28	3.80	1.09	0.52	9
Other Race/Ethnicity	1.11	1.50	1.85	3.28	1.53	0.55	15
Other Metro	1.15	1.39	1.71	2.90	1.45	0.44	22
Disadvantaged Urban	1.25	1.95	2.43	4.35	1.94	0.84	10
Advantaged Urban	1.21	1.54	1.92	3.97	1.63	0.66	12
Par. Ed. < HS	1.02	1.28	1.54	2.39	1.31	0.39	22
Par. Ed. = HS	1.08	1.35	1.56	2.80	1.40	0.39	25
Par. Ed. > HS	1.10	1.42	1.79	2.34	1.44	0.41	24
Par. Ed. = Coll.	1.13	1.26	1.43	2.65	1.29	0.30	37
Par. Ed. = IDK	1.06	1.27	1.45	3.16	1.36	0.47	16
Public School	1.23	1.41	1.67	2.53	1.46	0.35	35
Nonpublic School	1.10	1.64	2.24	4.21	1.72	0.83	8

* Distributions are based on 68 items.

Table D-10

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics
Across Cognitive Mathematics Cross-sectional Items
in the [Math-MainP] Samples

Grade 4 (Booklets 11-17)*

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.27	1.48	1.82	3.08	1.55	0.40	30
Male	1.22	1.45	1.70	2.75	1.48	0.36	34
Female	1.23	1.41	1.69	2.57	1.46	0.37	30
White	1.09	1.36	1.61	2.53	1.38	0.35	32
Black	1.11	1.28	1.59	2.69	1.36	0.38	25
Hispanic	1.16	1.41	1.63	2.63	1.43	0.37	29
Asian American	0.85	1.04	1.61	5.55	1.42	0.99	4
Other Race/Ethnicity	0.97	1.23	1.50	2.40	1.28	0.41	19
Other Metro	1.27	1.50	1.85	3.66	1.56	0.43	25
Disadvantaged Urban	1.20	1.37	1.66	2.87	1.46	0.39	27
Advantaged Urban	1.15	1.55	1.88	3.98	1.63	0.66	12
Par. Ed. < HS	0.97	1.14	1.40	3.21	1.24	0.39	20
Par. Ed. = HS	1.11	1.33	1.64	2.82	1.40	0.43	22
Par. Ed. > HS	1.06	1.25	1.51	2.23	1.30	0.30	38
Par. Ed. = Coll.	1.20	1.44	1.59	2.76	1.43	0.33	38
Par. Ed. = IDK	1.19	1.37	1.64	2.37	1.43	0.33	38
Public School	1.22	1.47	1.81	3.12	1.51	0.40	28
Nonpublic School	1.23	1.46	1.77	3.31	1.53	0.43	25

* Distributions are based on 109 items.

Table D-11

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics
Across Cognitive Mathematics Cross-sectional Items
in the [Math-MainP] Samples

Grade 8 (Booklets 8-14)*

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.31	1.59	1.91	4.20	1.68	0.54	19
Male	1.20	1.47	1.76	2.65	1.48	0.38	31
Female	1.20	1.43	1.76	2.57	1.49	0.41	26
White	1.23	1.55	1.87	3.99	1.61	0.54	18
Black	1.07	1.26	1.50	2.54	1.31	0.34	30
Hispanic	1.12	1.39	1.64	2.55	1.43	0.40	26
Asian American	1.14	1.44	1.85	4.18	1.55	0.57	15
Other Race/Ethnicity	1.00	1.30	1.88	5.19	1.58	0.87	7
Other Metro	1.29	1.53	1.70	3.65	1.56	0.46	23
Disadvantaged Urban	1.03	1.36	1.71	4.36	1.48	0.65	10
Advantaged Urban	1.12	1.86	3.43	9.20	2.51	1.94	3
Par. Ed. < HS	0.94	1.11	1.36	2.23	1.17	0.35	23
Par. Ed. = HS	1.03	1.22	1.42	2.65	1.26	0.33	28
Par. Ed. > HS	1.01	1.19	1.40	2.10	1.22	0.29	35
Par. Ed. = Coll.	1.25	1.50	1.83	2.86	1.56	0.44	25
Par. Ed. = IDK	1.09	1.33	1.56	2.09	1.33	0.32	34
Public School	1.24	1.56	1.90	4.53	1.64	0.57	16
Nonpublic School	1.23	1.45	1.86	3.88	1.61	0.57	16

* Distributions are based on 137 items.

Table D-12

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics
Across Cognitive Mathematics Cross-sectional Items
in the [Math-MainP] Samples

Grade 12 (Booklets 8-14)*

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.44	1.65	2.04	3.78	1.75	0.49	25
Male	1.26	1.56	1.84	2.55	1.56	0.40	31
Female	1.22	1.43	1.73	2.53	1.49	0.36	34
White	1.35	1.61	1.94	3.86	1.73	0.56	19
Black	0.99	1.22	1.44	2.38	1.26	0.33	29
Hispanic	1.03	1.31	1.66	3.10	1.40	0.50	16
Asian American	1.03	1.27	1.81	6.13	1.50	0.81	7
Other Race/Ethnicity	0.89	1.21	1.60	2.78	1.25	0.46	15
Other Metro	1.29	1.59	1.94	3.64	1.66	0.49	23
Disadvantaged Urban	1.24	1.84	2.85	6.68	2.13	1.20	6
Advantaged Urban	1.22	1.54	1.95	4.51	1.66	0.65	13
Par. Ed. < HS	1.00	1.22	1.40	2.22	1.22	0.31	31
Par. Ed. = HS	1.13	1.47	1.68	2.54	1.46	0.41	25
Par. Ed. > HS	1.11	1.31	1.54	2.42	1.35	0.35	30
Par. Ed. = Coll.	1.18	1.44	1.74	2.55	1.46	0.37	31
Par. Ed. = IDK	0.87	1.11	1.35	3.05	1.15	0.38	18
Public School	1.28	1.51	1.86	3.20	1.60	0.45	25
Nonpublic School	1.37	1.69	1.97	3.08	1.72	0.47	26

---

* Distributions are based on 145 items.

Table D-13

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics
Across Cognitive Mathematics Cross-sectional Items
in the [Math-MainT] Samples

Grade 4 (Booklet 28)*

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.49	1.71	2.00	3.66	1.83	0.56	21
Male	1.38	1.57	1.86	2.67	1.64	0.39	35
Female	1.29	1.59	1.89	2.61	1.59	0.42	28
White	1.35	1.51	1.94	3.07	1.63	0.49	22
Black	1.06	1.31	1.61	2.46	1.37	0.43	20
Hispanic	1.02	1.25	1.50	2.46	1.29	0.39	22
Asian American	1.04	1.38	1.99	3.38	1.55	0.67	10
Other Race/Ethnicity	0.80	0.92	1.23	2.09	1.04	0.35	18
Other Metro	1.39	1.75	2.02	3.89	1.79	0.60	17
Disadvantaged Urban	1.40	1.62	2.13	3.11	1.76	0.62	16
Advantaged Urban	1.15	1.51	2.25	6.73	1.83	1.16	5
Par. Ed. < HS	0.90	1.10	1.28	1.86	1.13	0.29	29
Par. Ed. = HS	1.00	1.26	1.63	2.44	1.34	0.46	17
Par. Ed. > HS	1.06	1.26	1.44	2.19	1.27	0.31	34
Par. Ed. = Coll.	1.26	1.49	1.70	2.48	1.53	0.36	36
Par. Ed. = IDK	1.08	1.31	1.50	2.94	1.32	0.36	26
Public School	1.45	1.63	1.95	3.73	1.77	0.58	18
Nonpublic School	1.53	2.06	2.76	5.33	2.28	1.02	10

* Distributions are based on 48 items.

Table D-14

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics
Across Cognitive Mathematics Cross-sectional Items
in the [Math-MainT] Samples

Grade 8 (Booklet 25)*

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.16	1.36	1.64	2.37	1.41	0.38	27
Male	1.03	1.22	1.50	2.20	1.26	0.33	28
Female	1.02	1.21	1.48	2.54	1.29	0.34	28
White	1.01	1.21	1.58	3.20	1.32	0.46	16
Black	1.05	1.36	1.66	2.62	1.37	0.42	21
Hispanic	0.99	1.19	1.43	2.24	1.23	0.32	28
Asian American	1.07	1.54	1.97	5.88	1.65	0.86	7
Other Race/Ethnicity	0.91	1.27	2.33	7.06	1.75	1.21	4
Other Metro	1.15	1.39	1.83	2.63	1.47	0.43	23
Disadvantaged Urban	1.07	1.50	1.90	3.92	1.63	0.71	10
Advantaged Urban	1.13	1.53	2.50	6.95	1.95	1.32	4
Par. Ed. < HS	0.99	1.24	1.50	2.52	1.28	0.36	25
Par. Ed. = HS	0.95	1.15	1.37	1.71	1.15	0.27	35
Par. Ed. > HS	1.02	1.22	1.34	1.82	1.20	0.23	54
Par. Ed. = Coll.	1.13	1.34	1.52	3.41	1.38	0.42	22
Par. Ed. = IDK	0.96	1.11	1.27	1.77	1.13	0.27	34
Public School	1.09	1.26	1.50	2.62	1.33	0.36	26
Nonpublic School	1.47	1.80	2.28	3.32	1.87	0.62	18

* Distributions are based on 75 items.

Table D-15

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics
Across Cognitive Mathematics Cross-sectional Items
in the [Math-MainT] Samples

Grade 12 (Booklet 25)*

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.39	1.70	2.03	3.82	1.74	0.52	22
Male	1.32	1.54	1.82	2.59	1.62	0.38	36
Female	1.24	1.54	1.84	3.12	1.63	0.51	20
White	1.32	1.62	1.93	3.97	1.69	0.54	20
Black	1.03	1.23	1.44	2.09	1.24	0.26	44
Hispanic	1.03	1.30	1.76	3.58	1.47	0.64	10
Asian American	1.31	1.68	2.23	4.96	1.92	0.89	9
Other Race/Ethnicity	1.02	1.45	1.82	2.63	1.47	0.46	20
Other Metro	1.30	1.46	1.92	3.44	1.63	0.50	21
Disadvantaged Urban	1.87	2.73	4.20	9.27	3.16	1.80	6
Advantaged Urban	1.46	1.90	2.54	4.36	2.07	0.85	12
Par. Ed. < HS	1.03	1.21	1.46	1.88	1.24	0.30	33
Par. Ed. = HS	1.05	1.23	1.43	2.04	1.26	0.31	33
Par. Ed. > HS	1.13	1.30	1.56	2.42	1.36	0.38	26
Par. Ed. = Coll.	1.26	1.48	1.77	2.93	1.55	0.41	28
Par. Ed. = IDK	0.97	1.17	1.39	1.82	1.18	0.30	30
Public School	1.24	1.44	1.85	3.52	1.56	0.49	20
Nonpublic School	1.81	2.21	2.79	6.59	2.38	0.99	11

* Distributions are based on 81 items.

Table D-16

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics
Across Cognitive Mathematics Trend Items
in the [RMS-Br86] and [MS-BrLT] Samples

Age 9 (Booklets 91-95)*

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.85	2.21	2.59	5.70	2.35	0.80	17
Male	1.53	1.82	2.19	3.04	1.87	0.48	30
Female	1.39	1.67	2.14	4.23	1.83	0.59	19
White	1.35	1.69	2.03	4.96	1.81	0.68	14
Black	1.29	1.66	2.18	4.54	1.82	0.70	13
Hispanic	1.29	1.55	2.00	4.33	1.67	0.58	17
Asian American	0.89	1.29	1.85	5.34	1.51	0.98	5
Other Race/Ethnicity	1.00	1.15	1.32	2.46	1.18	0.29	33
Other Metro	1.59	2.13	2.78	5.36	2.27	0.87	13
Disadvantaged Urban	1.56	2.28	3.46	13.36	2.97	2.28	3
Advantaged Urban	0.99	1.40	2.09	5.12	1.70	0.95	6
Par. Ed. < HS	0.97	1.15	1.42	2.94	1.21	0.37	22
Par. Ed. = HS	1.10	1.37	1.69	4.35	1.51	0.62	12
Par. Ed. > HS	1.08	1.28	1.51	2.17	1.33	0.33	32
Par. Ed. = Coll.	1.36	1.67	2.00	4.30	1.76	0.59	18
Par. Ed. = IDK	1.24	1.44	1.75	2.36	1.50	0.37	32
Public School	1.89	2.19	2.65	5.56	2.36	0.79	18
Nonpublic School	1.40	1.82	2.70	8.04	2.19	1.21	6

* Distributions are based on 117 items.

Table D-17

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics
Across Cognitive Mathematics Trend Items
in the [RMS-Br86] and [MS-BrLT] Samples

Age 13 (Booklets 91-95)*

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.60	2.06	2.78	5.07	2.27	0.91	12
Male	1.30	1.65	2.22	4.23	1.79	0.66	14
Female	1.36	1.66	2.17	3.32	1.74	0.56	19
White	1.32	1.75	2.31	4.33	1.93	0.84	10
Black	1.12	1.46	2.20	5.99	1.77	0.90	8
Hispanic	1.04	1.36	1.70	4.04	1.44	0.59	12
Asian American	0.97	1.40	2.61	12.58	2.14	1.92	2
Other Race/Ethnicity	0.82	1.04	1.29	2.89	1.08	0.34	20
Other Metro	1.50	1.98	2.68	5.26	2.21	0.96	11
Disadvantaged Urban	1.15	1.80	2.52	8.08	2.05	1.26	5
Advantaged Urban	1.07	1.64	2.33	6.60	1.89	1.17	5
Par. Ed. < HS	0.97	1.24	1.57	4.29	1.32	0.49	15
Par. Ed. = HS	1.17	1.39	1.69	3.75	1.45	0.44	21
Par. Ed. > HS	1.08	1.26	1.49	2.72	1.30	0.34	28
Par. Ed. = Coll.	1.25	1.61	2.14	4.94	1.78	0.71	12
Par. Ed. = IDK	1.03	1.22	1.44	2.75	1.27	0.37	24
Public School	1.57	2.04	2.82	5.03	2.27	0.94	12
Nonpublic School	0.88	1.43	2.23	5.06	1.65	0.98	6

* Distributions are based on 159 items.

Table D-18

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics
Across Cognitive Mathematics Trend Items
in the [RMS-Br86] Samples

Age 17 (Booklets 61-66)*

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.24	1.47	1.67	2.94	1.49	0.35	36
Male	1.18	1.41	1.65	2.54	1.44	0.37	30
Female	1.07	1.30	1.52	1.99	1.30	0.29	40
White	1.17	1.36	1.72	3.46	1.44	0.39	27
Black	1.13	1.44	1.84	3.72	1.54	0.57	14
Hispanic	0.93	1.21	1.55	3.75	1.29	0.53	12
Asian American	0.90	1.23	1.52	3.88	1.22	0.48	13
Other Race/Ethnicity	0.70	0.88	1.17	3.09	0.96	0.40	12
Other Metro	1.16	1.39	1.64	3.29	1.43	0.37	29
Disadvantaged Urban	1.30	1.81	2.49	3.90	1.91	0.80	11
Advantaged Urban	1.04	1.36	1.77	4.08	1.46	0.61	12
Par. Ed. < HS	1.07	1.30	1.67	2.98	1.37	0.42	21
Par. Ed. = HS	1.05	1.25	1.52	3.01	1.31	0.39	22
Par. Ed. > HS	1.12	1.28	1.54	3.19	1.39	0.41	22
Par. Ed. = Coll.	1.13	1.36	1.60	3.15	1.41	0.41	23
Par. Ed. = IDK	1.08	1.37	1.58	3.21	1.35	0.41	21
Public School	1.22	1.43	1.66	2.99	1.47	0.37	31
Nonpublic School	1.07	1.52	1.95	5.98	1.63	0.77	9

* Distributions are based on 164 items.

Table D-19

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics
Across Cognitive Mathematics Trend Items
in the [MS-Br86] and [MS-BrLT] Samples

Age 17 (Booklets 84-85 and 94-95)*

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.82	2.19	2.82	5.50	2.31	0.71	21
Male	1.51	1.83	2.20	3.68	1.90	0.55	24
Female	1.37	1.69	2.03	3.28	1.73	0.47	26
White	1.28	1.67	2.24	4.65	1.85	0.74	12
Black	1.39	2.08	2.93	8.34	2.34	1.29	7
Hispanic	1.10	1.50	2.12	4.89	1.71	0.84	8
Asian American	1.03	1.38	1.94	5.19	1.57	0.84	7
Other Race/Ethnicity	0.82	1.18	1.51	3.42	1.23	0.56	10
Other Metro	1.50	1.88	2.45	3.55	1.94	0.59	22
Disadvantaged Urban	1.65	2.81	4.30	12.48	3.40	2.43	4
Advantaged Urban	1.76	3.01	4.98	20.28	3.74	2.81	4
Par. Ed. < HS	0.97	1.23	1.49	3.54	1.28	0.42	18
Par. Ed. = HS	1.11	1.36	1.63	4.05	1.43	0.47	18
Par. Ed. > HS	1.16	1.37	1.57	4.16	1.40	0.42	23
Par. Ed. = Coll.	1.33	1.59	2.06	5.12	1.76	0.64	15
Par. Ed. = IDK	0.97	1.26	1.52	4.80	1.40	0.70	8
Public School	1.68	2.07	2.55	3.79	2.13	0.60	25
Nonpublic School	1.67	3.07	10.09	33.68	6.56	6.84	2

* Distributions are based on 161 items.

Table D-20

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics
Across Cognitive Science Cross-sectional Items
in the [Sci-MainP] Samples

Grade 4 (Booklets 18-24)*

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.22	1.48	1.75	2.81	1.52	0.38	31
Male	1.19	1.39	1.54	2.64	1.42	0.35	33
Female	1.20	1.42	1.59	2.34	1.42	0.32	40
White	1.14	1.39	1.64	2.56	1.43	0.39	27
Black	0.93	1.10	1.34	2.41	1.15	0.34	22
Hispanic	1.08	1.26	1.49	2.31	1.28	0.28	42
Asian American	0.90	1.21	1.96	6.83	1.62	1.10	4
Other Race/Ethnicity	1.01	1.21	1.43	2.91	1.25	0.36	24
Other Metro	1.20	1.51	1.74	2.99	1.52	0.41	27
Disadvantaged Urban	0.86	1.07	1.34	2.37	1.13	0.40	16
Advantaged Urban	1.07	1.38	1.62	2.59	1.39	0.42	22
Par. Ed. < HS	1.12	1.33	1.67	2.83	1.43	0.47	18
Par. Ed. = HS	1.09	1.28	1.52	2.37	1.31	0.33	30
Par. Ed. > HS	1.06	1.20	1.50	2.59	1.29	0.34	29
Par. Ed. = Coll.	1.18	1.38	1.60	3.35	1.40	0.35	32
Par. Ed. = IDK	1.16	1.36	1.58	2.34	1.37	0.31	40
Public School	1.20	1.42	1.64	2.88	1.46	0.36	33
Nonpublic School	1.13	1.34	1.66	2.30	1.41	0.34	34

* Distributions are based on 112 items.

Table D-21

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics
Across Cognitive Science Cross-sectional Items
in the [Sci-MainP] Samples

Grade 8 (Booklets 15-21)*

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.26	1.51	1.84	3.35	1.57	0.43	26
Male	1.14	1.40	1.64	2.36	1.42	0.32	38
Female	1.20	1.42	1.68	2.98	1.46	0.41	25
White	1.14	1.39	1.71	3.35	1.47	0.44	22
Black	1.09	1.33	1.65	2.74	1.39	0.42	22
Hispanic	1.04	1.22	1.57	2.24	1.31	0.37	25
Asian American	0.93	1.20	1.49	3.85	1.31	0.52	12
Other Race/Ethnicity	0.69	1.05	1.99	5.78	1.50	1.07	4
Other Metro	1.21	1.43	1.78	2.85	1.50	0.41	27
Disadvantaged Urban	1.07	1.54	1.94	4.55	1.62	0.70	11
Advantaged Urban	1.11	1.62	2.72	11.67	2.20	1.76	3
Par. Ed. < HS	1.06	1.26	1.53	2.15	1.29	0.33	31
Par. Ed. = HS	1.05	1.24	1.50	2.97	1.29	0.36	26
Par. Ed. > HS	1.06	1.20	1.45	2.51	1.27	0.32	31
Par. Ed. = Coll.	1.16	1.38	1.75	2.73	1.47	0.42	24
Par. Ed. = IDK	1.05	1.25	1.53	2.08	1.30	0.32	32
Public School	1.20	1.43	1.76	3.29	1.49	0.43	24
Nonpublic School	1.38	1.62	1.96	4.67	1.74	0.61	16

* Distributions are based on 146 items.

Table D-22

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics
Across Cognitive Science Cross-sectional Items
in the [Sci-MainP] Samples

Grade 12 (Booklets 15-21)*

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.31	1.53	1.78	2.88	1.56	0.38	33
Male	1.16	1.43	1.66	2.95	1.46	0.39	28
Female	1.17	1.34	1.58	2.39	1.39	0.31	39
White	1.22	1.49	1.75	3.59	1.51	0.42	26
Black	1.03	1.27	1.49	2.39	1.28	0.34	28
Hispanic	1.11	1.43	1.73	2.93	1.45	0.44	22
Asian American	1.01	1.30	2.29	10.35	1.90	1.58	3
Other Race/Ethnicity	0.84	1.15	1.49	4.90	1.24	0.58	9
Other Metro	1.31	1.52	1.78	2.75	1.56	0.37	35
Disadvantaged Urban	1.02	1.47	2.58	6.33	1.90	1.26	5
Advantaged Urban	1.06	1.44	2.03	3.33	1.57	0.67	11
Par. Ed. < HS	0.98	1.24	1.55	2.83	1.33	0.45	17
Par. Ed. = HS	1.09	1.27	1.50	2.48	1.31	0.35	27
Par. Ed. > HS	1.12	1.30	1.49	2.47	1.33	0.32	35
Par. Ed. = Coll.	1.16	1.34	1.57	2.37	1.38	0.32	37
Par. Ed. = IDK	0.99	1.29	1.60	3.69	1.36	0.50	15
Public School	1.21	1.42	1.65	2.55	1.45	0.34	35
Nonpublic School	1.36	1.76	2.11	3.55	1.81	0.58	19

---

* Distributions are based on 150 items.

Table D-23

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics
Across Cognitive Science Trend Items
in the [RMS-Br86] and [MS-BrLT] Samples

Age 9 (Booklets 91-95)*

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.58	1.88	2.23	3.55	1.95	0.53	27
Male	1.33	1.55	1.80	2.66	1.58	0.38	35
Female	1.44	1.62	1.96	3.54	1.72	0.46	28
White	1.23	1.45	1.71	3.39	1.51	0.44	24
Black	1.31	1.64	2.19	6.38	1.83	0.80	10
Hispanic	1.24	1.58	1.87	3.35	1.60	0.50	20
Asian American	0.99	1.39	1.95	8.25	1.59	1.02	5
Other Race/Ethnicity	1.00	1.17	1.34	1.84	1.19	0.25	45
Other Metro	1.53	1.88	2.24	3.70	1.90	0.57	22
Disadvantaged Urban	1.29	2.07	3.00	7.57	2.33	1.38	6
Advantaged Urban	1.05	1.49	2.09	4.02	1.63	0.78	9
Par. Ed. < HS	1.02	1.22	1.46	2.30	1.25	0.35	25
Par. Ed. = HS	1.10	1.32	1.55	3.26	1.37	0.40	23
Par. Ed. > HS	1.00	1.17	1.37	3.11	1.23	0.36	23
Par. Ed. = Coll.	1.28	1.51	1.83	3.29	1.61	0.46	24
Par. Ed. = IDK	1.21	1.39	1.64	2.84	1.45	0.37	30
Public School	1.59	1.89	2.22	3.32	1.93	0.51	28
Nonpublic School	1.39	1.90	2.62	7.06	2.09	1.06	8

* Distributions are based on 129 items.

Table D-24

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics
Across Cognitive Science Trend Items
in the [RMS-Br86] and [MS-BrLT] Samples

Age 13 (Booklets 91-95)*

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.55	1.88	2.41	4.35	2.04	0.67	19
Male	1.26	1.57	1.93	3.86	1.65	0.53	19
Female	1.39	1.64	1.92	3.41	1.69	0.46	26
White	1.29	1.59	1.98	4.21	1.68	0.56	18
Black	1.07	1.45	1.97	5.79	1.66	0.84	8
Hispanic	1.04	1.31	1.80	4.23	1.45	0.56	13
Asian American	1.01	1.51	2.36	9.58	1.99	1.56	3
Other Race/Ethnicity	0.86	1.02	1.29	3.08	1.11	0.38	17
Other Metro	1.43	1.75	2.21	4.38	1.92	0.66	17
Disadvantaged Urban	1.03	1.76	2.65	6.67	2.00	1.22	5
Advantaged Urban	0.92	1.63	2.41	5.82	1.79	1.01	6
Par. Ed. < HS	0.96	1.18	1.46	2.58	1.23	0.37	22
Par. Ed. = HS	1.09	1.31	1.54	2.27	1.33	0.32	34
Par. Ed. > HS	1.08	1.28	1.53	2.32	1.32	0.33	32
Par. Ed. = Coll.	1.28	1.59	1.86	2.98	1.62	0.47	23
Par. Ed. = IDK	1.09	1.35	1.61	2.71	1.38	0.38	26
Public School	1.57	1.94	2.47	4.58	2.11	0.73	17
Nonpublic School	0.95	1.32	1.82	3.81	1.46	0.71	9

* Distributions are based on 160 items.

Table D-25

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics
Across Cognitive Science Trend Items
in the [RMS-Br86] Samples

Age 17 (Booklets 61-66)*

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.14	1.36	1.57	2.25	1.38	0.33	35
Male	1.15	1.34	1.49	2.61	1.38	0.33	35
Female	1.12	1.30	1.49	2.42	1.33	0.31	36
White	1.05	1.30	1.54	2.70	1.34	0.38	25
Black	1.10	1.44	1.76	3.85	1.53	0.61	12
Hispanic	0.98	1.17	1.58	3.31	1.30	0.49	14
Asian American	0.85	1.05	1.29	3.60	1.16	0.55	9
Other Race/Ethnicity	0.87	1.30	1.67	3.33	1.30	0.61	9
Other Metro	1.13	1.33	1.62	2.51	1.38	0.36	28
Disadvantaged Urban	1.09	1.63	2.09	4.09	1.69	0.71	11
Advantaged Urban	0.95	1.17	1.59	3.70	1.32	0.57	11
Par. Ed. < HS	1.03	1.31	1.59	4.11	1.39	0.53	14
Par. Ed. = HS	1.16	1.38	1.70	2.49	1.44	0.40	26
Par. Ed. > HS	1.14	1.38	1.68	2.90	1.43	0.40	26
Par. Ed. = Coll.	1.07	1.33	1.56	2.23	1.34	0.34	31
Par. Ed. = IDK	1.09	1.30	1.66	2.55	1.37	0.41	22
Public School	1.12	1.34	1.60	2.43	1.36	0.34	32
Nonpublic School	0.97	1.35	1.83	4.01	1.44	0.63	10

* Distributions are based on 121 items.

Table D-26

Distributions of Design Effects by Demographic Subgroups
for Proportion-correct Statistics
Across Cognitive Science Trend Items
in the [MS-Br86] and [MS-BrLT] Samples

Age 17 (Booklets 84-85 and 94-95)*

Subgroup	LoQ	Median	HiQ	Max.	Mean	Standard Deviation	Degrees of Freedom
Total	1.63	2.03	2.53	8.29	2.19	0.89	12
Male	1.28	1.54	1.98	5.60	1.66	0.61	15
Female	1.36	1.67	2.05	5.63	1.79	0.62	17
White	1.24	1.57	1.93	9.40	1.78	1.01	6
Black	1.10	1.55	2.01	7.85	1.72	0.94	7
Hispanic	0.99	1.35	1.88	5.70	1.53	0.77	8
Asian American	1.05	1.54	2.23	5.60	1.78	1.00	6
Other Race/Ethnicity	0.88	1.09	1.49	2.75	1.17	0.40	17
Other Metro	1.41	1.73	2.11	7.99	1.90	0.85	10
Disadvantaged Urban	1.08	1.85	3.20	12.86	2.43	1.98	3
Advantaged Urban	1.24	1.90	3.13	16.99	2.79	2.62	2
Par. Ed. < HS	1.06	1.28	1.50	4.47	1.35	0.49	15
Par. Ed. = HS	1.12	1.32	1.61	3.46	1.39	0.40	24
Par. Ed. > HS	1.11	1.27	1.58	3.36	1.38	0.42	21
Par. Ed. = Coll.	1.23	1.53	1.82	5.83	1.63	0.66	12
Par. Ed. = IDK	0.95	1.23	1.54	2.87	1.31	0.49	14
Public School	1.55	1.99	2.51	8.11	2.16	0.90	11
Nonpublic School	1.12	1.59	2.57	14.04	2.40	2.41	2

* Distributions are based on 160 items.

Table D-27

Distributions of Design Effects by Demographic Subgroups
for Average Writing Scores
Across Cognitive Writing Trend Items
in the [RW-Br84] Samples

Grade 4 (Booklets 51-56)*

Subgroup	Median	Max.
Total	1.56	1.87
Male	1.27	1.72
Female	1.56	1.95
White	1.22	1.98
Black	1.67	2.82
Hispanic	1.12	2.14
Asian American	1.06	1.62
Other Race/Ethnicity	1.02	1.11
Other Metro	1.29	1.77
Disadvantaged Urban	2.02	4.73
Advantaged Urban	1.77	2.36
Par. Ed. < HS	1.28	1.69
Par. Ed. = HS	1.31	1.49
Par. Ed. > HS	1.29	1.61
Par. Ed. = Coll.	1.37	1.57
Par. Ed. = IDK	1.21	1.37
Public School	1.62	1.89
Nonpublic School	1.20	1.55

* Distributions are based on 6 items.

Table D-28

Distributions of Design Effects by Demographic Subgroups
for Average Writing Scores
Across Cognitive Writing Trend Items
in the [RW-Br84] Samples

Grade 8 (Booklets 51-56)*

Subgroup	Median	Max.
Total	1.43	2.18
Male	1.24	1.89
Female	1.21	1.44
White	1.49	1.55
Black	1.15	1.35
Hispanic	1.19	1.53
Asian American	1.84	3.29
Other Race/Ethnicity	0.97	1.52
Other Metro	1.53	2.07
Disadvantaged Urban	1.42	2.06
Advantaged Urban	1.35	1.92
Par. Ed. < HS	1.19	1.63
Par. Ed. = HS	1.44	1.56
Par. Ed. > HS	1.28	1.55
Par. Ed. = Coll.	1.17	1.90
Par. Ed. = IDK	1.06	1.61
Public School	1.45	1.96
Nonpublic School	1.38	1.57

* Distributions are based on 6 items.

Table D-29

Distributions of Design Effects by Demographic Subgroups
for Average Writing Scores
Across Cognitive Writing Trend Items
in the [RW-Br84] Samples

Grade 11 (Booklets 51-56)*

Subgroup	Median	Max.
Total	1.28	1.80
Male	1.45	1.82
Female	1.13	1.51
White	1.34	1.73
Black	0.94	0.99
Hispanic	0.99	1.67
Asian American	1.24	1.97
Other Race/Ethnicity	1.28	1.65
Other Metro	1.17	2.08
Disadvantaged Urban	1.31	1.62
Advantaged Urban	1.62	2.26
Par. Ed. < HS	1.22	2.01
Par. Ed. = HS	1.35	1.63
Par. Ed. > HS	1.39	1.65
Par. Ed. = Coll.	1.12	2.06
Par. Ed. = IDK	1.11	2.25
Public School	1.31	1.63
Nonpublic School	2.30	3.22

---

* Distributions are based on 6 items.

APPENDIX E

Subscale Item Counts and IRT Parameters

Appendix E

SUBSCALE ITEM COUNTS AND IRT PARAMETERS

Tables E-1 through E-7 in this appendix show, for each cross-sectional block and booklet, the number of items that were used in a scale or subscale for the reading, mathematics, and science cross-sectional assessments.

This appendix also contains 23 tables of IRT (item response theory) parameters for NAEP items that were scaled in each subject area and study (cross-sectional or trend) for which IRT scales or subscales were created.

For each NAEP item used in scaling, the tables show the corresponding IRT parameters (A, B, and C) and standard errors (S.E.), the block in which the item appears for each age class (BLOCK), and the position of the item within the block (ITEM).

IRT parameters for items used in cross-sectional scales are shown for reading in Tables E-11 through E-14, for mathematics in Tables E-18 through E-23, and for science in Tables E-27 through E-30. IRT parameters for items used in trend scales are shown for reading in Tables E-8 through E-10, for mathematics in Tables E-15 through E-17, and for science in Tables E-24 through E-26. The standard errors of items for which parameters were fixed are listed as (0.000).

Note that item parameters shown in this appendix are in the metrics used for the original calibration of the scale. The transformations needed to represent these parameters in terms of the metric of the final reporting scales are given in Chapters 12 through 14.

Table E-1

Number of Reading Items Used in Cross-sectional Scaling
for Each Age/Grade, by Block

Block	Age 9/Grade 4	Age 13/Grade 8	Age 17/Grade 12
RC	7	9	18
RD	7	10	15
RE	12	19	19
RF	9	14	14
RG	12	15	14
RH	11	15	17
RI	9	14	14
Total	67	96	111

Table E-2

Number of Reading Items Used in Cross-sectional Scaling
for Each Age/Grade, by Booklet

Booklet	Age 9/Grade 4	Age 13/Grade 8	Age 17/Grade 12
1	23	33	47
2	31	44	48
3	32	48	50
4	30	43	42
5	30	39	49
6	27	39	46
7	28	42	51

Table E-3

Number of Mathematics Items Used in Cross-sectional Scaling for Each Age/Grade and Subscale, by Block

Block	Numbers and Operations	Measurement	Geometry	Data Analysis, Statistics, and Probability	Algebra and Functions	Total
Age 9/Grade 4						
MC	9	5	1	2	2	19
MD	7	2	2	1	2	14
ME	4	1	3	1	2	11
MF	6	4	3	1	3	17
MG	8	4	2	0	1	15
MH	9	2	2	0	1	14
MI	9	2	1	0	3	15
Total	52	20	14	5	14	105
Age 13/Grade 8						
MC	9	4	3	4	3	23
MD	7	4	4	2	4	21
ME	3	1	6	3	3	16
MF	7	5	3	3	3	21
MG	3	3	5	3	4	18
MH	8	2	2	3	3	18
MI	9	1	3	1	5	19
Total	46	20	26	19	25	136
Age 17/Grade 12						
MC	7	4	5	3	4	23
MD	7	4	3	2	6	22
ME	3	1	3	5	5	17
MF	5	5	2	3	4	19
MG	3	3	6	2	7	21
MH	5	3	3	4	6	21
MI	7	2	3	3	5	20
Total	37	22	25	22	37	143

Table E-4

Number of Mathematics Items Used in Cross-sectional Scaling
for Each Age/Grade and Subscale, by Booklet

Booklet	Numbers and Operations	Measurement	Geometry	Data Analysis, Statistics, and Probability	Algebra and Functions	Total
			Age 9/Grade 4			
11	22	11	6	4	7	50
12	19	7	7	2	5	40
13	19	7	8	2	6	42
14	23	10	6	1	7	47
15	26	11	5	2	4	48
16	25	6	5	1	6	43
17	22	8	5	3	7	45
			Age 13/Grade 8			
8	23	13	10	9	10	65
9	13	8	15	8	11	55
10	18	8	11	9	9	55
11	19	9	11	7	12	58
12	20	9	10	10	10	59
13	24	7	9	6	12	58
14	21	6	12	8	11	58
			Age 17/Grade 12			
8	19	13	10	8	14	64
9	13	8	12	9	18	60
10	13	9	8	12	15	57
11	15	10	11	8	16	60
12	15	10	14	9	17	65
13	19	9	9	9	17	63
14	17	7	11	11	14	60

Table E-5

Number of Mathematics Estimation Items Used in Cross-sectional Scaling
for Each Age/Grade, by Block

Block	Age 9/Grade 4	Age 13/Grade 8	Age 17/Grade 12
MJ	19	21	21
MK	0	24	24
Total	19	45	45

Table E-6

Number of Science Items Used in Cross-sectional Scaling
for Each Age/Grade and Subscale, by Block

Block	Life Sciences	Physical Sciences	Earth and Space Sciences	Nature of Science	Total
Age 9/Grade 4					
SC	5	6	4	2	17
SD	6	3	8	3	20
SE	5	5	7	2	19
SF	4	4	2	2	12
SG	5	6	5	4	20
SH	2	6	2	0	10
SI	3	5	3	1	12
Total	30	35	31	14	110
Age 13/Grade 8					
SC	9	5	4	4	22
SD	9	5	8	4	26
SE	6	6	10	3	25
SF	4	7	3	3	17
SG	5	5	8	4	22
SH	4	6	4	0	14
SI	5	5	4	2	16
Total	42	39	41	20	142
Age 17/Grade 12					
SC	8	8	5	4	25
SD	9	7	8	5	29
SE	5	9	5	2	21
SF	5	9	2	0	16
SG	6	6	8	5	25
SH	4	7	5	0	16
SI	4	2	2	2	10
Total	41	48	35	18	142

Table E-7

Number of Science Items Used in Cross-sectional Scaling
for Each Age/Grade and Subscale, by Booklet

Booklet	Life Sciences	Physical Sciences	Earth and Space Sciences	Nature of Science	Total
Age 9/Grade 4					
18	15	13	14	7	49
19	16	14	20	9	59
20	11	15	11	4	41
21	12	15	10	7	44
22	12	18	11	6	47
23	11	14	13	4	42
24	13	16	14	5	48
Age 13/Grade 8					
15	22	17	15	11	65
16	20	16	26	11	73
17	14	19	17	6	56
18	14	17	15	9	55
19	18	16	16	8	58
20	18	16	16	6	56
21	20	16	18	9	63
Age 17/Grade 12					
15	22	24	15	9	70
16	20	22	21	12	75
17	14	25	12	2	53
18	15	17	12	7	51
19	18	21	18	9	66
20	17	16	15	7	55
21	17	19	12	8	56

Table E-8
1990 IRT Parameters, Reading Trend Sample, Age 9

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 9 Block	Age 9 Item
N001101	0.412	(0.119)	0.737	(0.499)	0.277	(0.059)	H	5
N001501	1.813	(0.000)	-1.607	(0.000)	0.214	(0.000)	H	10
N001502	1.658	(0.000)	-0.482	(0.000)	0.201	(0.000)	H	11
N001503	1.300	(0.000)	-1.144	(0.000)	0.211	(0.000)	H	12
N001504	1.648	(0.000)	-0.565	(0.000)	0.238	(0.000)	H	13
N001601	0.820	(0.000)	-0.621	(0.365)	0.210	(0.000)	J	12
N001602	1.089	(0.000)	-0.253	(0.127)	0.254	(0.000)	J	13
N001603	0.659	(0.207)	0.473	(0.339)	0.291	(0.056)	J	14
N001604	0.774	(0.000)	0.628	(0.000)	0.208	(0.000)	J	15
N001802	1.040	(0.000)	1.775	(0.000)	0.220	(0.000)	J	20
N002001	1.042	(0.000)	0.479	(0.000)	0.202	(0.000)	K	9
N002002	1.408	(0.500)	0.305	(0.154)	0.214	(0.039)	K	10
N002003	1.405	(0.000)	-0.053	(0.057)	0.224	(0.000)	K	11
N002101	0.973	(0.000)	1.443	(0.000)	0.240	(0.000)	K	18
N002102	0.789	(0.000)	2.358	(0.000)	0.163	(0.000)	K	19
N002401	1.047	(0.000)	0.066	(0.000)	0.156	(0.000)	L	22
N002702	1.280	(0.000)	0.268	(0.000)	0.187	(0.000)	L	20
N002801	1.750	(0.000)	-0.613	(0.000)	0.239	(0.000)	L	24
N002802	1.554	(0.000)	-0.751	(0.000)	0.209	(0.000)	L	25
N002804	0.451	(0.000)	2.086	(0.000)	0.000	(0.000)	L	26
N003001	0.719	(0.000)	2.213	(0.000)	0.189	(0.000)	M	10
N003002	0.411	(0.106)	0.177	(0.387)	0.228	(0.058)	M	11
N003101	0.787	(0.184)	-0.410	(0.206)	0.238	(0.055)	M	14
N003102	1.240	(0.000)	0.083	(0.000)	0.184	(0.000)	M	15
N003104	0.483	(0.000)	3.114	(0.000)	0.000	(0.000)	M	16
N003701	0.862	(0.000)	-1.063	(0.357)	0.255	(0.000)	N	23
N003702	1.397	(0.000)	-0.134	(0.084)	0.262	(0.000)	N	24
N003704	0.635	(0.000)	0.901	(0.000)	0.000	(0.000)	N	25
N003801	1.318	(0.000)	1.203	(0.000)	0.289	(0.000)	O	12
N003802	0.559	(0.000)	-0.309	(0.831)	0.203	(0.000)	O	13
N003803	0.702	(0.000)	2.737	(0.000)	0.202	(0.000)	O	14
N004101	0.874	(0.000)	-1.120	(0.635)	0.242	(0.000)	O	17
N004201	0.83[illegible]	(0.000)	0.487	(0.000)	0.219	(0.000)	O	18
N004202	0.5[illegible]	(0.000)	0.950	(0.000)	0.300	(0.000)	O	19
N004701	1.2[illegible]5	(0.000)	-0.323	(0.135)	0.200	(0.000)	Q	10
N004702	0.545	(0.000)	-0.771	(0.948)	0.246	(0.000)	Q	11
N004703	0.845	(0.000)	-0.351	(0.232)	0.227	(0.000)	Q	12
N004801	1.032	(0.000)	-1.261	(0.000)	0.233	(0.000)	Q	13
N004901	1.134	(0.000)	0.435	(0.000)	0.245	(0.000)	Q	14
N005101	0.562	(0.000)	-2.865	(6.061)	0.230	(0.000)	Q	15
N008601	1.386	(0.000)	-1.018	(0.000)	0.200	(0.000)	H	6
N008602	1.032	(0.000)	-0.741	(0.155)	0.212	(0.000)	H	7
N008603	1.081	(0.000)	-1.033	(0.073)	0.188	(0.000)	H	8
N008701	0.878	(0.000)	-3.145	(3.425)	0.229	(0.000)	H	9
N008801	1.539	(0.000)	-1.600	(0.000)	0.240	(0.000)	J	18
N008901	1.247	(0.000)	-1.012	(0.000)	0.213	(0.000)	J	21
N008902	1.013	(0.000)	-1.188	(0.203)	0.203	(0.000)	J	22
N009001	1.068	(0.000)	-0.077	(0.092)	0.206	(0.000)	K	12
N009002	1.102	(0.357)	0.364	(0.198)	0.246	(0.046)	K	13
N009003	1.549	(0.674)	0.566	(0.180)	0.218	(0.036)	K	14
N009004	0.962	(0.000)	-0.443	(0.195)	0.173	(0.000)	K	15
N009101	0.896	(0.000)	-1.525	(0.475)	0.216	(0.000)	K	16
N009201	1.418	(0.000)	-1.532	(0.000)	0.242	(0.000)	K	17
N009401	1.207	(0.000)	-1.588	(0.237)	0.153	(0.000)	L	23

Table E-8 (continued)
1990 IRT Parameters, Reading Trend Sample, Age 9

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 9 Block	Item
N009601	0.808	(0.000)	-2.370	(1.671)	0.186	(0.000)	L	21
N009701	0.949	(0.000)	-0.514	(0.195)	0.241	(0.000)	M	5
N009702	1.804	(0.000)	-0.527	(0.000)	0.285	(0.000)	M	6
N009703	1.150	(0.000)	0.077	(0.000)	0.255	(0.000)	M	7
N009704	0.888	(0.000)	0.113	(0.000)	0.169	(0.000)	M	8
N009705	1.132	(0.000)	-0.792	(0.104)	0.195	(0.000)	M	9
N009801	1.224	(0.000)	-2.454	(1.006)	0.236	(0.000)	N	12
N009901	0.842	(0.000)	-0.993	(0.272)	0.246	(0.000)	N	13
N010002	0.977	(0.000)	-1.075	(0.209)	0.195	(0.000)	N	18
N010003	1.352	(0.000)	-0.788	(0.000)	0.245	(0.000)	N	19
N010102	0.828	(0.000)	0.106	(0.000)	0.234	(0.000)	N	21
N010103	1.789	(0.000)	-0.894	(0.000)	0.219	(0.000)	N	22
N010201	0.984	(0.000)	-2.129	(1.109)	0.234	(0.000)	O	16
N010301	0.527	(0.000)	-2.397	(4.097)	0.234	(0.000)	O	15
N010401	0.693	(0.000)	-1.251	(1.349)	0.224	(0.000)	O	20
N010402	0.990	(0.000)	0.476	(0.000)	0.244	(0.000)	O	21
N010403	1.225	(0.000)	0.896	(0.000)	0.252	(0.000)	O	22
N010801	0.917	(0.000)	-0.473	(0.242)	0.209	(0.000)	Q	16
N010902	1.273	(0.000)	-0.168	(0.100)	0.269	(0.000)	Q	18
N010903	1.229	(0.300)	-0.644	(0.160)	0.241	(0.056)	Q	19
N010904	1.282	(0.000)	-0.139	(0.094)	0.259	(0.000)	Q	20
N011001	1.004	(0.000)	-0.916	(0.112)	0.194	(0.000)	R	5
N011002	1.427	(0.000)	0.006	(0.000)	0.279	(0.000)	R	6
N011003	1.935	(0.000)	-0.779	(0.000)	0.245	(0.000)	R	7
N011004	1.512	(0.000)	-0.409	(0.000)	0.232	(0.000)	R	8
N011101	1.187	(0.000)	-0.411	(0.056)	0.181	(0.000)	R	9
N011201	0.736	(0.000)	-0.102	(0.129)	0.253	(0.000)	R	10
N011301	1.687	(0.000)	-0.501	(0.000)	0.247	(0.000)	R	11
N011302	0.812	(0.000)	-0.265	(0.173)	0.240	(0.000)	R	12
N011401	0.855	(0.000)	0.866	(0.000)	0.327	(0.000)	R	13
N011402	0.573	(0.000)	0.564	(0.000)	0.278	(0.000)	R	14
N011403	1.110	(0.000)	0.917	(0.000)	0.282	(0.000)	R	15
N011404	1.154	(0.000)	0.764	(0.000)	0.207	(0.000)	R	16
N013201	1.566	(0.000)	-0.468	(0.000)	0.244	(0.000)	V	29
N013301	0.949	(0.000)	-1.698	(0.105)	0.242	(0.000)	V	30
N013401	1.390	(0.000)	-0.127	(0.072)	0.171	(0.000)	V	31
N013402	1.252	(0.000)	-0.643	(0.000)	0.209	(0.000)	V	32
N013403	1.213	(0.000)	-0.022	(0.038)	0.215	(0.000)	V	33
N014001	1.045	(0.000)	-0.989	(0.081)	0.236	(0.000)	M	13
N014101	0.562	(0.000)	-1.134	(1.266)	0.221	(0.000)	Q	21
N014201	0.867	(0.000)	-1.205	(0.281)	0.175	(0.000)	V	34
N014301	1.515	(0.000)	-0.585	(0.000)	0.207	(0.000)	N	14
N014302	0.885	(0.210)	-0.309	(0.176)	0.217	(0.052)	N	15
N014303	1.160	(0.000)	-0.937	(0.065)	0.186	(0.000)	N	16
N014501	0.588	(0.000)	-1.734	(0.000)	0.000	(0.000)	V	35
N014502	0.493	(0.000)	-1.888	(0.000)	0.000	(0.000)	V	35
N014503	0.999	(0.000)	-2.014	(0.000)	0.000	(0.000)	V	35

Table E-9
1990 IRT Parameters, Reading Trend Sample, Age 13

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 13 Block	Age 13 Item
N001101	0.339	(0.072)	0.865	(0.331)	0.252	(0.057)	H	6
N001201	0.715	(0.190)	1.209	(0.196)	0.388	(0.045)	H	7
N001202	0.713	(0.000)	1.260	(0.000)	0.265	(0.000)	H	8
N001301	0.625	(0.151)	0.684	(0.260)	0.508	(0.051)	H	9
N001302	0.473	(0.000)	-2.012	(0.000)	0.573	(0.000)	H	10
N001303	1.057	(0.000)	0.936	(0.000)	0.280	(0.000)	H	11
N001401	0.711	(0.000)	0.240	(0.000)	0.230	(0.000)	H	12
N001501	1.658	(0.000)	-1.572	(0.000)	0.212	(0.000)	H	13
N001502	1.684	(0.000)	-0.474	(0.000)	0.166	(0.000)	H	14
N001503	1.334	(0.000)	-0.868	(0.000)	0.220	(0.000)	H	15
N001504	1.519	(0.000)	-0.626	(0.000)	0.215	(0.000)	H	16
N001601	0.361	(0.000)	-1.300	(0.000)	0.239	(0.000)	J	11
N001602	0.590	(0.000)	-1.528	(0.000)	0.254	(0.000)	J	12
N001603	0.658	(0.000)	0.201	(0.000)	0.256	(0.000)	J	13
N001604	0.714	(0.000)	-0.219	(0.000)	0.244	(0.000)	J	14
N001701	0.831	(0.127)	-0.804	(0.186)	0.239	(0.057)	J	17
N001702	0.341	(0.000)	4.420	(0.000)	0.245	(0.000)	J	18
N001703	0.630	(0.000)	-0.173	(0.000)	0.227	(0.000)	J	19
N001802	0.777	(0.000)	1.148	(0.000)	0.221	(0.000)	J	21
N001901	0.879	(0.000)	0.362	(0.000)	0.256	(0.000)	J	22
N002001	0.803	(0.000)	0.020	(0.000)	0.193	(0.000)	K	9
N002002	0.936	(0.000)	-0.187	(0.000)	0.190	(0.000)	K	10
N002003	1.154	(0.000)	-0.280	(0.000)	0.228	(0.000)	K	11
N002101	0.899	(0.199)	1.293	(0.114)	0.181	(0.032)	K	12
N002102	1.233	(0.000)	1.202	(0.000)	0.162	(0.000)	K	13
N002201	1.080	(0.000)	-0.005	(0.000)	0.236	(0.000)	K	14
N002202	0.751	(0.000)	-0.671	(0.000)	0.272	(0.000)	K	15
N002203	0.561	(0.000)	-1.698	(0.000)	0.238	(0.000)	K	16
N002401	0.824	(0.000)	-0.877	(0.000)	0.147	(0.000)	L	22
N002501	0.467	(0.000)	0.434	(0.000)	0.210	(0.000)	L	23
N002701	0.702	(0.129)	0.654	(0.136)	0.209	(0.042)	L	24
N002801	1.879	(0.000)	-0.661	(0.000)	0.268	(0.000)	L	25
N002802	0.924	(0.000)	-1.365	(0.000)	0.213	(0.000)	L	26
N002902	0.418	(0.000)	-1.153	(0.000)	0.245	(0.000)	M	6
N002903	1.601	(0.000)	-0.455	(0.000)	0.250	(0.000)	M	7
N002904	1.122	(0.000)	0.218	(0.000)	0.245	(0.000)	M	8
N002905	0.648	(0.000)	0.674	(0.000)	0.201	(0.000)	M	9
N002906	1.730	(0.000)	-0.289	(0.000)	0.294	(0.000)	M	10
N003001	0.542	(0.000)	1.816	(0.000)	0.165	(0.000)	M	11
N003002	0.298	(0.000)	0.041	(0.000)	0.159	(0.000)	M	12
N003003	1.132	(0.000)	3.004	(0.000)	0.090	(0.000)	M	13
N003101	1.059	(0.000)	-0.934	(0.000)	0.234	(0.000)	M	14
N003102	1.270	(0.000)	-0.306	(0.000)	0.213	(0.000)	M	15
N003201	0.875	(0.000)	-0.634	(0.000)	0.251	(0.000)	N	12
N003202	1.066	(0.000)	0.301	(0.000)	0.188	(0.000)	N	13
N003203	1.044	(0.000)	0.184	(0.000)	0.210	(0.000)	N	14
N003204	0.686	(0.000)	0.865	(0.000)	0.259	(0.000)	N	15
N003301	0.851	(0.000)	-0.295	(0.000)	0.208	(0.000)	N	16
N003401	0.892	(0.000)	-0.161	(0.000)	0.152	(0.000)	N	17
N003501	0.773	(0.000)	-0.100	(0.000)	0.230	(0.000)	N	18
N003601	0.790	(0.000)	-1.239	(0.000)	0.232	(0.000)	N	19
N003602	0.884	(0.000)	-0.097	(0.000)	0.223	(0.000)	N	20
N003701	0.962	(0.000)	-0.454	(0.000)	0.232	(0.000)	N	21
N003702	0.758	(0.000)	0.069	(0.000)	0.237	(0.000)	N	22

Table E-9 (continued)
1990 IRT Parameters, Reading Trend Sample, Age 13

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 13 Block	Age 13 Item
N003704	0.592	(0.075)	0.384	(0.075)	0.000	(0.000)	N	23
N003801	0.513	(0.140)	1.705	(0.288)	0.191	(0.046)	O	12
N003802	0.231	(0.000)	-1.483	(0.000)	0.165	(0.000)	O	13
N003803	0.344	(0.000)	3.638	(0.000)	0.233	(0.000)	O	14
N003901	1.031	(0.000)	-2.151	(0.000)	0.232	(0.000)	O	16
N004002	0.424	(0.000)	-1.964	(0.000)	0.243	(0.000)	O	15
N004101	0.984	(0.000)	-1.219	(0.000)	0.234	(0.000)	O	17
N004201	0.670	(0.000)	0.193	(0.000)	0.238	(0.000)	O	18
N004202	0.723	(0.000)	0.163	(0.000)	0.219	(0.000)	O	19
N004301	1.123	(0.329)	0.694	(0.117)	0.248	(0.045)	O	20
N004303	0.408	(0.061)	0.858	(0.130)	0.000	(0.000)	O	21
N004401	0.832	(0.000)	-2.323	(0.000)	0.233	(0.000)	P	7
N004402	0.865	(0.000)	0.164	(0.000)	0.273	(0.000)	P	8
N004403	1.115	(0.000)	-1.525	(0.000)	0.220	(0.000)	P	9
N004501	0.489	(0.087)	0.454	(0.220)	0.231	(0.054)	P	10
N004502	0.511	(0.000)	-0.963	(0.000)	0.232	(0.000)	P	11
N004601	1.263	(0.274)	0.638	(0.095)	0.266	(0.038)	P	12
N004602	1.267	(0.000)	0.193	(0.000)	0.247	(0.000)	P	13
N004603	1.330	(0.000)	-0.347	(0.000)	0.252	(0.000)	P	14
N004605	0.700	(0.077)	-0.806	(0.115)	0.000	(0.000)	P	15
N004701	1.190	(0.000)	-0.716	(0.000)	0.205	(0.000)	Q	7
N004702	0.505	(0.000)	-1.297	(0.000)	0.233	(0.000)	Q	8
N004703	0.667	(0.000)	-1.028	(0.000)	0.229	(0.000)	Q	9
N004801	0.972	(0.000)	-1.500	(0.000)	0.227	(0.000)	Q	10
N004901	0.679	(0.000)	0.120	(0.000)	0.215	(0.000)	Q	11
N005001	1.089	(0.000)	2.283	(0.000)	0.214	(0.000)	Q	13
N005002	0.498	(0.000)	2.342	(0.000)	0.234	(0.000)	Q	14
N005003	0.632	(0.000)	2.298	(0.000)	0.105	(0.000)	Q	15
N005101	0.488	(0.000)	-3.052	(0.000)	0.237	(0.000)	Q	12
N005201	0.401	(0.000)	2.055	(0.000)	0.611	(0.000)	Q	16
N005202	0.367	(0.000)	0.984	(0.000)	0.221	(0.000)	Q	17
N005203	0.524	(0.148)	2.129	(0.335)	0.236	(0.039)	Q	18
N005301	0.906	(0.000)	0.173	(0.000)	0.276	(0.000)	Q	19
N005302	1.192	(0.000)	0.807	(0.000)	0.178	(0.000)	Q	20
N005303	0.755	(0.145)	0.861	(0.138)	0.206	(0.043)	Q	21
N005304	1.654	(0.000)	0.288	(0.000)	0.166	(0.000)	Q	22
N005305	0.776	(0.000)	-0.676	(0.000)	0.249	(0.000)	Q	23
N005403	0.936	(0.116)	-0.581	(0.143)	0.229	(0.050)	R	7
N005404	1.054	(0.000)	-1.168	(0.000)	0.233	(0.000)	R	8
N005405	1.156	(0.000)	0.188	(0.000)	0.228	(0.000)	R	9
N005406	1.044	(0.141)	-0.267	(0.120)	0.257	(0.046)	R	10
N005407	1.253	(0.000)	-0.368	(0.000)	0.296	(0.000)	R	11
N005503	0.648	(0.000)	0.272	(0.000)	0.217	(0.000)	R	14
N005504	1.260	(0.000)	1.210	(0.000)	0.200	(0.000)	R	15
N005505	0.831	(0.000)	-0.868	(0.000)	0.236	(0.000)	R	16
N005601	1.521	(0.000)	-0.360	(0.000)	0.262	(0.000)	R	17
N005602	1.608	(0.000)	0.694	(0.000)	0.239	(0.000)	R	18
N005603	0.963	(0.000)	-0.373	(0.000)	0.228	(0.000)	R	19

Table E-10
1990 IRT Parameters, Reading Trend Sample, Age 17

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 17 Block	Item
N001301	0.923	(0.000)	0.862	(0.000)	0.564	(0.000)	H	10
N001302	0.964	(0.000)	-1.521	(0.000)	0.558	(0.000)	H	11
N001303	0.798	(0.000)	0.397	(0.000)	0.277	(0.000)	H	12
N001401	0.768	(0.000)	0.155	(0.000)	0.283	(0.000)	H	13
N001501	1.419	(0.000)	-1.425	(0.000)	0.249	(0.000)	H	14
N001502	0.873	(0.113)	-0.436	(0.175)	0.245	(0.059)	H	15
N001503	1.147	(0.000)	-0.777	(0.000)	0.237	(0.000)	H	16
N001504	0.962	(0.116)	-0.483	(0.168)	0.237	(0.057)	H	17
N001507	0.395	(0.065)	3.211	(0.344)	0.000	(0.000)	H	19
N001701	0.458	(0.000)	-0.954	(0.000)	0.257	(0.000)	J	12
N001702	0.650	(0.217)	3.231	(0.447)	0.252	(0.032)	J	13
N001703	0.990	(0.148)	0.289	(0.150)	0.259	(0.058)	J	14
N001901	0.926	(0.000)	-0.298	(0.000)	0.260	(0.000)	J	15
N001904	0.690	(0.085)	-0.208	(0.133)	0.000	(0.000)	J	17
N002001	1.152	(0.000)	0.353	(0.000)	0.222	(0.000)	K	9
N002002	1.151	(0.000)	0.084	(0.000)	0.227	(0.000)	K	10
N002003	1.719	(0.369)	-0.001	(0.118)	0.252	(0.055)	K	11
N002101	0.437	(0.000)	1.264	(0.000)	0.139	(0.000)	K	12
N002102	1.110	(0.000)	1.034	(0.000)	0.142	(0.000)	K	13
N002201	1.232	(0.162)	-0.328	(0.146)	0.244	(0.057)	K	14
N002202	1.005	(0.000)	-0.392	(0.000)	0.244	(0.000)	K	15
N002203	0.467	(0.000)	-2.206	(0.000)	0.257	(0.000)	K	16
N002501	0.431	(0.000)	0.123	(0.000)	0.261	(0.000)	L	27
N002701	0.767	(0.000)	0.903	(0.000)	0.140	(0.000)	L	28
N002702	0.829	(0.000)	-0.059	(0.000)	0.134	(0.000)	L	29
N002801	1.862	(0.000)	-0.889	(0.000)	0.245	(0.000)	L	30
N002802	1.155	(0.000)	-1.191	(0.000)	0.247	(0.000)	L	31
N002902	0.564	(0.000)	-0.782	(0.000)	0.255	(0.000)	M	6
N002903	2.251	(0.000)	-0.218	(0.000)	0.253	(0.000)	M	7
N002904	0.874	(0.121)	-0.274	(0.188)	0.250	(0.059)	M	8
N002905	0.707	(0.000)	0.899	(0.000)	0.215	(0.000)	M	9
N002906	1.529	(0.000)	-0.330	(0.000)	0.223	(0.000)	M	10
N003001	0.633	(0.106)	1.429	(0.143)	0.143	(0.045)	M	11
N003002	0.382	(0.000)	0.874	(0.000)	0.144	(0.000)	M	12
N003003	1.528	(0.000)	2.201	(0.000)	0.085	(0.000)	M	13
N003101	0.912	(0.000)	-1.092	(0.000)	0.256	(0.000)	M	14
N003102	1.320	(0.000)	-0.345	(0.000)	0.254	(0.000)	M	15
N003104	0.761	(0.100)	2.197	(0.113)	0.000	(0.000)	M	16
N003201	1.215	(0.000)	-0.683	(0.000)	0.249	(0.000)	N	21
N003202	0.958	(0.000)	-0.070	(0.000)	0.252	(0.000)	N	22
N003203	1.141	(0.000)	0.522	(0.000)	0.224	(0.000)	N	23
N003204	1.027	(0.000)	-0.275	(0.000)	0.235	(0.000)	N	24
N003301	1.560	(0.000)	-0.141	(0.000)	0.262	(0.000)	N	25
N003501	0.889	(0.115)	-0.309	(0.180)	0.249	(0.059)	N	27
N003601	1.121	(0.000)	-0.850	(0.000)	0.243	(0.000)	N	28
N003602	1.182	(0.000)	-0.203	(0.000)	0.262	(0.000)	N	29
N003701	1.101	(0.000)	-0.057	(0.000)	0.277	(0.000)	N	30
N003702	0.784	(0.000)	0.039	(0.000)	0.244	(0.000)	N	31
N003704	0.673	(0.074)	0.619	(0.076)	0.000	(0.000)	N	32
N003801	0.880	(0.155)	1.296	(0.111)	0.143	(0.042)	O	12
N003802	0.216	(0.000)	-1.821	(0.000)	0.156	(0.000)	O	13
N003803	0.578	(0.000)	2.532	(0.000)	0.259	(0.000)	O	14
N004201	0.958	(0.000)	0.319	(0.000)	0.257	(0.000)	O	21
N004202	0.502	(0.000)	0.554	(0.000)	0.263	(0.000)	O	22

Table E-10 (continued)
1990 IRT Parameters, Reading Trend Sample, Age 17

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 17 Block	Age 17 Item
N004301	0.698	(0.000)	0.281	(0.000)	0.260	(0.000)	O	23
N004303	0.436	(0.066)	1.085	(0.113)	0.000	(0.000)	O	24
N004501	0.593	(0.099)	0.603	(0.212)	0.264	(0.060)	P	20
N004502	0.499	(0.000)	-1.079	(0.000)	0.256	(0.000)	P	21
N004601	0.975	(0.000)	0.671	(0.000)	0.215	(0.000)	P	22
N004602	0.939	(0.138)	0.087	(0.160)	0.245	(0.057)	P	23
N004603	1.290	(0.000)	-0.177	(0.000)	0.253	(0.000)	P	24
N004605	0.567	(0.079)	-0.206	(0.177)	0.000	(0.000)	P	25
N004901	0.774	(0.000)	0.318	(0.000)	0.255	(0.000)	Q	10
N005001	1.448	(0.000)	1.879	(0.000)	0.238	(0.000)	Q	7
N005002	0.793	(0.182)	2.172	(0.168)	0.261	(0.040)	Q	8
N005003	0.758	(0.176)	2.619	(0.190)	0.123	(0.034)	Q	9
N005201	0.737	(0.000)	1.363	(0.000)	0.587	(0.000)	Q	11
N005202	0.383	(0.067)	0.675	(0.302)	0.263	(0.060)	Q	12
N005203	0.755	(0.000)	2.154	(0.000)	0.303	(0.000)	Q	13
N005503	0.791	(0.000)	0.662	(0.000)	0.250	(0.000)	R	14
N005504	1.257	(0.000)	0.851	(0.000)	0.193	(0.000)	R	15
N005505	1.100	(0.000)	-0.804	(0.000)	0.249	(0.000)	R	16
N015101	0.696	(0.000)	0.548	(0.000)	0.218	(0.000)	R	17
N015102	2.297	(0.000)	0.778	(0.000)	0.198	(0.000)	R	18
N015103	2.013	(0.000)	0.875	(0.000)	0.175	(0.000)	R	19
N015104	1.681	(0.000)	0.581	(0.000)	0.247	(0.000)	R	20
N015201	0.917	(0.121)	-0.629	(0.192)	0.251	(0.060)	N	26
N015502	1.398	(0.000)	0.441	(0.000)	0.274	(0.000)	P	16
N015503	0.681	(0.000)	1.012	(0.000)	0.283	(0.000)	P	17
N015504	1.382	(0.000)	0.409	(0.000)	0.272	(0.000)	P	18
N015505	0.495	(0.000)	-0.057	(0.000)	0.271	(0.000)	P	19
N015901	0.874	(0.000)	0.714	(0.000)	0.304	(0.000)	Q	14
N015902	1.210	(0.000)	1.015	(0.000)	0.254	(0.000)	Q	15
N015903	1.342	(0.302)	1.516	(0.094)	0.236	(0.038)	Q	16
N015905	0.645	(0.075)	1.322	(0.080)	0.000	(0.000)	Q	17
N016001	0.801	(0.000)	-0.044	(0.000)	0.253	(0.000)	O	15
N016002	0.717	(0.000)	1.291	(0.000)	0.277	(0.000)	O	16
N016003	0.730	(0.000)	0.770	(0.000)	0.282	(0.000)	O	17
N016004	1.114	(0.194)	0.179	(0.152)	0.259	(0.060)	O	18
N016005	1.515	(0.000)	0.440	(0.000)	0.266	(0.000)	O	19
N016006	0.811	(0.000)	0.731	(0.000)	0.227	(0.000)	O	20
N017001	1.253	(0.000)	0.811	(0.000)	0.314	(0.000)	H	7
N017002	1.224	(0.000)	1.355	(0.000)	0.219	(0.000)	H	8
N017003	0.862	(0.213)	2.230	(0.146)	0.205	(0.036)	H	9

Table E-11
1990 IRT Parameters, Reading Answer Booklet Bridge Sample, Age 9/Grade 4

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 9/Grade 4 Block	Item
N001501	2.292	(0.000)	-0.586	(0.000)	0.252	(0.000)	R3	4
N001502	1.860	(0.000)	0.293	(0.000)	0.207	(0.000)	R3	5
N001503	1.559	(0.000)	-0.163	(0.000)	0.261	(0.000)	R3	6
N001504	1.802	(0.000)	0.212	(0.000)	0.276	(0.000)	R3	7
N001601	0.690	(0.000)	-0.124	(0.000)	0.197	(0.000)	R4	4
N001602	0.981	(0.000)	0.377	(0.000)	0.232	(0.000)	R4	5
N001603	0.784	(0.000)	0.996	(0.000)	0.272	(0.000)	R4	6
N001604	0.773	(0.000)	1.224	(0.000)	0.256	(0.000)	R4	7
N002401	1.000	(0.000)	0.587	(0.000)	0.119	(0.000)	R4	1
N002801	1.549	(0.000)	-0.053	(0.000)	0.131	(0.000)	R3	2
N002802	1.544	(0.000)	-0.181	(0.000)	0.172	(0.000)	R3	3
N003701	0.899	(0.000)	-0.369	(0.000)	0.269	(0.000)	R4	2
N003702	1.061	(0.000)	0.204	(0.000)	0.270	(0.000)	R4	3
N004801	0.746	(0.000)	-0.797	(0.000)	0.276	(0.000)	R3	1
N005101	0.649	(0.000)	-2.380	(0.000)	0.242	(0.000)	R2	2
N008701	0.575	(0.000)	-2.938	(0.000)	0.255	(0.000)	R2	1
N010501	1.585	(0.000)	-0.816	(0.000)	0.225	(0.000)	R2	8
N010502	0.881	(0.000)	-0.568	(0.000)	0.211	(0.000)	R2	9
N010503	1.420	(0.000)	-0.762	(0.000)	0.302	(0.000)	R2	10
N010504	1.589	(0.000)	-0.309	(0.000)	0.151	(0.000)	R2	11
N013101	1.550	(0.000)	-0.791	(0.000)	0.311	(0.000)	R2	4
N013102	0.973	(0.000)	0.078	(0.000)	0.189	(0.000)	R2	5
N013103	0.740	(0.000)	-0.053	(0.000)	0.157	(0.000)	R2	6
N013104	0.473	(0.000)	0.495	(0.000)	0.221	(0.000)	R2	7
N013301	1.165	(0.000)	-1.335	(0.000)	0.240	(0.000)	R2	3
N014301	1.502	(0.000)	0.123	(0.000)	0.189	(0.000)	R2	12
N014302	1.030	(0.000)	0.369	(0.000)	0.205	(0.000)	R2	13
N014303	1.765	(0.000)	-0.358	(0.000)	0.271	(0.000)	R2	14
N021101	0.678	(0.000)	-1.780	(0.000)	0.223	(0.000)	R7	4
N021102	0.931	(0.000)	-0.712	(0.000)	0.189	(0.000)	R7	5
N021103	0.880	(0.000)	-0.786	(0.000)	0.171	(0.000)	R7	6
N021401	0.978	(0.000)	0.027	(0.000)	0.206	(0.000)	R7	12
N021402	1.448	(0.000)	-0.196	(0.000)	0.178	(0.000)	R7	13
N021403	1.443	(0.000)	-1.064	(0.000)	0.208	(0.000)	R7	14
N021404	1.343	(0.000)	0.459	(0.000)	0.203	(0.000)	R7	15
R000301	1.430	(0.000)	0.398	(0.000)	0.215	(0.000)	R6	1
R000302	1.237	(0.000)	-0.200	(0.000)	0.141	(0.000)	R6	2
R000303	0.976	(0.000)	-0.013	(0.000)	0.174	(0.000)	R6	3
R000304	0.993	(0.000)	-0.215	(0.000)	0.216	(0.000)	R6	4
R000401	1.237	(0.000)	0.343	(0.000)	0.134	(0.000)	R6	5
R000402	0.750	(0.000)	-1.308	(0.000)	0.203	(0.000)	R6	6
R000403	0.832	(0.000)	0.952	(0.000)	0.266	(0.000)	R6	7
R000404	1.196	(0.000)	-0.595	(0.000)	0.160	(0.000)	R6	8
R000501	1.532	(0.000)	-0.377	(0.000)	0.226	(0.000)	R6	9
R000502	1.173	(0.000)	0.070	(0.000)	0.240	(0.000)	R6	10
R000503	0.576	(0.000)	1.685	(0.000)	0.130	(0.000)	R6	11
R000504	0.793	(0.000)	0.071	(0.000)	0.278	(0.000)	R6	12
R000505	0.944	(0.000)	0.076	(0.000)	0.196	(0.000)	R6	13
R000601	1.047	(0.000)	-0.828	(0.000)	0.244	(0.000)	R7	1
R000602	0.716	(0.000)	-0.429	(0.000)	0.216	(0.000)	R7	2
R000603	1.011	(0.000)	-0.827	(0.000)	0.170	(0.000)	R7	3
R000701	0.907	(0.000)	0.227	(0.000)	0.156	(0.000)	R7	7
R000702	1.127	(0.000)	0.151	(0.000)	0.194	(0.000)	R7	8
R000703	1.267	(0.000)	-0.881	(0.000)	0.238	(0.000)	R7	9

Table E-11 (continued)
1990 IRT Parameters, Reading Answer Booklet Bridge Sample, Age 9/Grade 4

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 9/Grade 4 Block	Item
R000704	0.590	(0.000)	-1.498	(0.000)	0.221	(0.000)	R7	10
R000705	0.508	(0.000)	-0.324	(0.000)	0.211	(0.000)	R7	11
R000801	0.658	(0.000)	-1.663	(0.000)	0.199	(0.000)	R5	6
R000802	1.055	(0.000)	-1.130	(0.000)	0.185	(0.000)	R5	7
R000803	0.857	(0.000)	-0.161	(0.000)	0.149	(0.000)	R5	8
R000804	1.335	(0.000)	-0.170	(0.000)	0.096	(0.000)	R5	9
R000805	1.368	(0.000)	-0.530	(0.000)	0.228	(0.000)	R5	10
R000806	1.544	(0.000)	-0.570	(0.000)	0.552	(0.000)	R5	11
R000807	1.125	(0.000)	-0.598	(0.000)	0.000	(0.000)	R5	11
R000901	1.161	(0.000)	-0.306	(0.000)	0.211	(0.000)	R5	1
R000902	0.901	(0.000)	-0.407	(0.000)	0.243	(0.000)	R5	2
R000903	1.508	(0.000)	-0.778	(0.000)	0.184	(0.000)	R5	3
R000904	1.660	(0.000)	1.137	(0.000)	0.097	(0.000)	R5	4
R000905	1.330	(0.053)	-0.566	(0.037)	0.111	(0.020)	R5	5

Table E-12
1990 IRT Parameters, Reading Cross-sectional Sample, Age 9/Grade 4

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 9/Grade 4 Block	Item
N001501	1.621	(0.074)	-0.489	(0.040)	0.203	(0.024)	R3	4
N001502	1.654	(0.087)	0.276	(0.024)	0.169	(0.012)	R3	5
N001503	1.614	(0.085)	-0.054	(0.034)	0.268	(0.017)	R3	6
N001504	1.757	(0.099)	0.186	(0.028)	0.271	(0.013)	R3	7
N001601	0.616	(0.035)	-0.208	(0.104)	0.153	(0.037)	R4	4
N001602	1.015	(0.065)	0.349	(0.048)	0.226	(0.020)	R4	5
N001603	0.699	(0.065)	1.076	(0.063)	0.228	(0.021)	R4	6
N001604	0.774	(0.067)	1.143	(0.050)	0.207	(0.018)	R4	7
N002401	0.913	(0.057)	0.758	(0.038)	0.122	(0.016)	R4	1
N002801	1.629	(0.073)	0.111	(0.025)	0.144	(0.013)	R3	2
N002802	1.373	(0.063)	-0.096	(0.035)	0.173	(0.018)	R3	3
N003701	0.907	(0.053)	-0.441	(0.090)	0.254	(0.038)	R4	2
N003702	0.974	(0.059)	0.170	(0.056)	0.223	(0.023)	R4	3
N004801	0.604	(0.035)	-0.973	(0.155)	0.231	(0.055)	R3	1
R000801	0.816	(0.033)	-1.136	(0.085)	0.171	(0.041)	R5	6
R000802	0.951	(0.038)	-0.900	(0.067)	0.155	(0.035)	R5	7
R000803	1.002	(0.046)	-0.030	(0.045)	0.164	(0.020)	R5	8
R000804	1.791	(0.070)	0.015	(0.020)	0.138	(0.011)	R5	9
R000805	1.365	(0.064)	-0.399	(0.040)	0.183	(0.022)	R5	10
R000806	1.419	(0.094)	-0.508	(0.064)	0.476	(0.025)	R5	11
R000807	1.009	(0.027)	-0.237	(0.019)	0.000	(0.000)	R5	11
R000901	1.171	(0.053)	-0.340	(0.044)	0.164	(0.022)	R5	1
R000902	0.784	(0.038)	-0.388	(0.073)	0.154	(0.031)	R5	2
R000903	1.531	(0.060)	-0.659	(0.035)	0.123	(0.023)	R5	3
R000904	1.574	(0.102)	1.213	(0.023)	0.128	(0.006)	R5	4
R000905	1.365	(0.051)	-0.323	(0.031)	0.128	(0.017)	R5	5
R010301	0.767	(0.023)	1.066	(0.027)	0.000	(0.000)	R6	1
R010401	1.293	(0.069)	-0.736	(0.065)	0.235	(0.037)	R6	2
R010402	1.331	(0.066)	0.411	(0.031)	0.231	(0.013)	R6	3
R010403	1.570	(0.088)	0.534	(0.024)	0.214	(0.011)	R6	4
R010404	0.873	(0.051)	0.341	(0.052)	0.170	(0.021)	R6	5
R010501	0.394	(0.097)	4.248	(0.599)	0.152	(0.016)	R6	6
R010502	1.564	(0.086)	-0.178	(0.038)	0.249	(0.019)	R6	7
R010503	1.243	(0.071)	0.412	(0.035)	0.221	(0.015)	R6	8
R010504	1.193	(0.079)	0.688	(0.033)	0.224	(0.013)	R6	9
R010601	1.016	(0.055)	-0.311	(0.063)	0.212	(0.029)	R7	1
R010602	0.986	(0.053)	-0.065	(0.058)	0.239	(0.024)	R7	2
R010603	0.837	(0.052)	0.509	(0.049)	0.173	(0.019)	R7	3
R010701	0.786	(0.068)	1.283	(0.047)	0.214	(0.015)	R7	4
R010702	0.684	(0.048)	0.393	(0.077)	0.187	(0.027)	R7	5
R010703	1.463	(0.080)	0.199	(0.032)	0.262	(0.014)	R7	6
R010704	1.258	(0.055)	0.034	(0.033)	0.142	(0.016)	R7	7
R010705	0.889	(0.048)	-0.029	(0.061)	0.176	(0.026)	R7	8
R010706	1.020	(0.058)	0.221	(0.049)	0.239	(0.019)	R7	9
R010707	0.905	(0.055)	0.309	(0.053)	0.206	(0.021)	R7	10
R010708	1.372	(0.074)	-0.106	(0.039)	0.205	(0.019)	R7	11
R010709	0.841	(0.048)	0.045	(0.066)	0.193	(0.026)	R7	12
R011001	1.259	(0.061)	0.227	(0.031)	0.134	(0.015)	R8	1
R011002	0.663	(0.053)	0.878	(0.065)	0.191	(0.023)	R8	2
R011003	1.246	(0.171)	1.730	(0.064)	0.388	(0.009)	R8	3
R011004	0.850	(0.086)	1.610	(0.057)	0.244	(0.012)	R8	4
R011005	1.048	(0.081)	1.050	(0.036)	0.252	(0.012)	R8	5
R011006	0.946	(0.069)	1.255	(0.038)	0.195	(0.011)	R8	6
R011101	1.229	(0.127)	1.543	(0.046)	0.311	(0.009)	R8	7

Table E-12 (continued)
1990 IRT Parameters, Reading Cross-sectional Sample, Age 9/Grade 4

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 9/Grade 4 Block	Age 9/Grade 4 Item
R011102	2.238	(0.499)	2.308	(0.080)	0.152	(0.004)	R8	8
R011103	0.956	(0.114)	1.737	(0.067)	0.284	(0.011)	R8	9
R011104	0.705	(0.071)	1.281	(0.060)	0.233	(0.019)	R8	10
R011105	1.464	(0.354)	2.684	(0.164)	0.243	(0.006)	R8	11
R011501	0.469	(0.127)	4.033	(0.616)	0.206	(0.013)	R9	1
R011502	0.655	(0.059)	1.022	(0.064)	0.209	(0.022)	R9	2
R011503	1.167	(0.057)	0.095	(0.039)	0.178	(0.018)	R9	3
R011504	1.492	(0.083)	-0.280	(0.044)	0.275	(0.022)	R9	4
R011505	1.116	(0.058)	0.371	(0.034)	0.158	(0.015)	R9	5
R011506	0.549	(0.060)	1.868	(0.079)	0.129	(0.018)	R9	6
R011507	0.663	(0.044)	0.503	(0.073)	0.170	(0.025)	R9	7
R011508	1.793	(0.098)	0.164	(0.027)	0.259	(0.013)	R9	8
R011509	0.564	(0.023)	1.773	(0.060)	0.000	(0.000)	R9	9

Table E-13
1990 IRT Parameters, Reading Cross-sectional Sample, Age 13/Grade 8

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 13/Grade 8 Block	Item
N001501	1.559	(0.000)	-1.253	(0.000)	0.149	(0.000)	R3	4
N001502	1.718	(0.000)	-0.525	(0.000)	0.174	(0.000)	R3	5
N001503	1.621	(0.000)	-0.845	(0.000)	0.238	(0.000)	R3	6
N001504	1.473	(0.000)	-0.592	(0.000)	0.271	(0.000)	R3	7
N001601	0.457	(0.000)	-1.086	(0.000)	0.241	(0.000)	R4	4
N001602	1.032	(0.000)	-0.821	(0.000)	0.248	(0.000)	R4	5
N001603	0.836	(0.000)	-0.035	(0.000)	0.250	(0.000)	R4	6
N001604	0.855	(0.000)	-0.350	(0.000)	0.182	(0.000)	R4	7
N002001	1.369	(0.000)	0.017	(0.000)	0.151	(0.000)	R4	8
N002002	1.119	(0.000)	-0.190	(0.000)	0.154	(0.000)	R4	9
N002003	1.374	(0.000)	-0.283	(0.000)	0.259	(0.000)	R4	10
N002101	0.876	(0.000)	0.971	(0.000)	0.222	(0.000)	R6	13
N002102	1.259	(0.000)	1.094	(0.000)	0.168	(0.000)	R6	14
N002401	1.027	(0.000)	-0.667	(0.000)	0.171	(0.000)	R4	1
N002801	1.507	(0.000)	-0.886	(0.000)	0.168	(0.000)	R3	2
N002802	1.421	(0.000)	-1.084	(0.000)	0.226	(0.000)	R3	3
N003001	0.955	(0.000)	1.253	(0.000)	0.184	(0.000)	R6	10
N003002	0.378	(0.000)	0.763	(0.000)	0.149	(0.000)	R6	11
N003003	1.908	(0.000)	2.208	(0.000)	0.094	(0.000)	R6	12
N003101	1.199	(0.000)	-0.703	(0.000)	0.202	(0.000)	R3	8
N003102	1.267	(0.000)	-0.504	(0.000)	0.149	(0.000)	R3	9
N003201	0.969	(0.000)	-0.735	(0.000)	0.162	(0.000)	R6	1
N003202	1.147	(0.000)	0.039	(0.000)	0.243	(0.000)	R6	2
N003203	1.014	(0.000)	0.302	(0.000)	0.247	(0.000)	R6	3
N003204	0.879	(0.000)	0.549	(0.000)	0.287	(0.000)	R6	4
N003701	0.801	(0.000)	-1.172	(0.000)	0.172	(0.000)	R4	2
N003702	0.856	(0.000)	-0.321	(0.000)	0.245	(0.000)	R4	3
N004201	1.082	(0.000)	-0.017	(0.000)	0.251	(0.000)	R6	5
N004202	0.611	(0.000)	0.343	(0.000)	0.319	(0.000)	R6	6
N004801	0.962	(0.000)	-1.516	(0.000)	0.230	(0.000)	R3	1
N005001	1.856	(0.000)	1.756	(0.000)	0.240	(0.000)	R6	7
N005002	0.885	(0.000)	2.017	(0.000)	0.339	(0.000)	R6	8
N005003	0.827	(0.000)	2.092	(0.000)	0.164	(0.000)	R6	9
R001501	0.708	(0.000)	-1.283	(0.000)	0.188	(0.000)	R5	1
R001502	0.543	(0.000)	-0.156	(0.000)	0.108	(0.000)	R5	2
R001503	0.985	(0.000)	-1.148	(0.000)	0.205	(0.000)	R5	3
R001504	1.271	(0.000)	-0.304	(0.000)	0.282	(0.000)	R5	4
R001601	1.580	(0.000)	-0.985	(0.000)	0.144	(0.000)	R5	5
R001602	1.145	(0.000)	-0.644	(0.000)	0.209	(0.000)	R5	6
R001603	0.893	(0.000)	-0.458	(0.000)	0.179	(0.000)	R5	7
R001604	0.694	(0.000)	-1.220	(0.000)	0.183	(0.000)	R5	8
R001605	0.796	(0.000)	0.257	(0.000)	0.170	(0.000)	R5	9
R001701	0.959	(0.000)	-0.047	(0.000)	0.275	(0.000)	R5	10
R001702	1.650	(0.000)	-0.787	(0.000)	0.234	(0.000)	R5	11
R001703	0.499	(0.000)	1.022	(0.000)	0.126	(0.000)	R5	12
R001704	0.705	(0.000)	0.479	(0.000)	0.171	(0.000)	R5	13
R001801	1.230	(0.000)	-0.195	(0.000)	0.225	(0.000)	R5	14
R001802	1.744	(0.000)	0.247	(0.000)	0.135	(0.000)	R5	15
R001803	0.812	(0.000)	0.874	(0.000)	0.243	(0.000)	R5	16
R001804	1.882	(0.000)	0.163	(0.000)	0.290	(0.000)	R5	17
R001805	1.586	(0.000)	0.280	(0.000)	0.269	(0.000)	R5	18
R001806	0.771	(0.000)	-0.073	(0.000)	0.172	(0.000)	R5	19
R010001	2.199	(0.128)	0.513	(0.018)	0.306	(0.008)	R7	9
R010002	1.165	(0.059)	0.195	(0.037)	0.315	(0.015)	R7	10

Table E-13 (continued)
1990 IRT Parameters, Reading Cross-sectional Sample, Age 13/Grade 8

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 13/Grade 8 Block	Item
R010003	1.402	(0.067)	0.268	(0.027)	0.257	(0.012)	R7	11
R010004	1.975	(0.104)	-0.087	(0.023)	0.289	(0.012)	R7	12
R010005	1.123	(0.048)	0.093	(0.032)	0.166	(0.015)	R7	13
R010006	1.724	(0.095)	0.553	(0.023)	0.320	(0.009)	R7	14
R010007	0.851	(0.021)	0.312	(0.018)	0.000	(0.000)	R7	15
R010601	1.009	(0.043)	-1.470	(0.084)	0.229	(0.04[illegible])	R7	1
R010602	0.784	(0.041)	-1.034	(0.116)	0.327	(0.046)	R7	2
R010603	1.019	(0.043)	-0.479	(0.051)	0.212	(0.025)	R7	3
R010801	0.973	(0.038)	-0.310	(0.045)	0.153	(0.021)	R7	4
R010802	0.850	(0.050)	0.497	(0.047)	0.259	(0.017)	R7	5
R010803	0.646	(0.036)	-0.079	(0.088)	0.218	(0.030)	R7	6
R010804	0.848	(0.045)	0.192	(0.053)	0.266	(0.020)	R7	7
R010805	0.514	(0.014)	0.180	(0.023)	0.000	(0.000)	R7	8
R011001	1.018	(0.042)	-0.429	(0.047)	0.163	(0.024)	R8	1
R011002	0.561	(0.043)	0.897	(0.071)	0.183	(0.024)	R8	2
R011003	0.611	(0.041)	0.354	(0.084)	0.230	(0.027)	R8	3
R011004	0.657	(0.049)	0.871	(0.058)	0.227	(0.020)	R8	4
R011005	0.925	(0.053)	0.411	(0.044)	0.287	(0.016)	R8	5
R011006	0.957	(0.058)	0.549	(0.040)	0.303	(0.015)	R8	6
R011101	1.336	(0.065)	0.527	(0.024)	0.244	(0.010)	R9	11
R011102	1.644	(0.093)	0.995	(0.020)	0.214	(0.007)	R9	12
R011103	0.935	(0.053)	0.600	(0.037)	0.236	(0.014)	R9	13
R011104	0.939	(0.043)	0.102	(0.041)	0.181	(0.018)	R9	14
R011105	0.669	(0.078)	2.059	(0.091)	0.263	(0.013)	R9	15
R011201	0.891	(0.0[illegible]9)	-1.272	(0.090)	0.225	(0.046)	R8	7
R011202	1.360	(0.055)	-0.544	(0.040)	0.275	(0.021)	R8	8
R011203	0.612	(0.061)	1.646	(0.066)	0.241	(0.016)	R8	9
R011204	0.594	(0.051)	1.345	(0.058)	0.200	(0.018)	R8	10
R011301	1.194	(0.052)	0.562	(0.024)	0.177	(0.010)	R8	11
R011302	1.290	(0.057)	-0.153	(0.034)	0.244	(0.017)	R8	12
R011303	0.684	(0.073)	1.604	(0.066)	0.309	(0.015)	R8	13
R011304	1.496	(0.065)	0.118	(0.025)	0.241	(0.012)	R8	14
R011305	1.509	(0.072)	0.610	(0.021)	0.225	(0.009)	R8	15
R011602	0.652	(0.039)	0.386	(0.066)	0.174	(0.024)	R9	2
R011603	0.594	(0.036)	-0.365	(0.117)	0.223	(0.038)	R9	3
R011604	1.044	(0.050)	0.217	(0.037)	0.252	(0.015)	R9	4
R011605	1.268	(0.045)	-0.270	(0.031)	0.174	(0.016)	R9	5
R011606	1.245	(0.083)	1.198	(0.029)	0.282	(0.008)	R9	6
R011607	0.847	(0.039)	-0.265	(0.060)	0.196	(0.025)	R9	7
R011701	0.798	(0.037)	0.091	(0.050)	0.159	(0.020)	R9	8
R011702	0.849	(0.037)	-0.661	(0.075)	0.228	(0.032)	R9	9
R011703	0.573	(0.034)	0.412	(0.076)	0.150	(0.025)	R9	10

Table E-14
1990 IRT Parameters, Reading Cross-sectional Sample, Age 17/Grade 12

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 17/Grade 12 Block	Item
N002101	0.616	(0.000)	0.189	(0.000)	0.218	(0.000)	R6	13
N002102	1.183	(0.000)	0.177	(0.000)	0.192	(0.000)	R6	14
N003001	0.890	(0.000)	0.440	(0.000)	0.179	(0.000)	R6	10
N003002	0.377	(0.000)	0.238	(0.000)	0.154	(0.000)	R6	11
N003003	1.465	(0.000)	1.204	(0.000)	0.103	(0.000)	R6	12
N003201	0.930	(0.000)	-1.676	(0.000)	0.212	(0.000)	R6	1
N003202	0.966	(0.000)	-0.964	(0.000)	0.340	(0.000)	R6	2
N003203	0.798	(0.000)	-0.452	(0.000)	0.231	(0.000)	R6	3
N003204	0.949	(0.000)	-1.268	(0.000)	0.292	(0.000)	R6	4
N004201	0.977	(0.000)	-0.638	(0.000)	0.274	(0.000)	R6	5
N004202	0.613	(0.000)	-0.387	(0.000)	0.341	(0.000)	R6	6
N005001	2.144	(0.000)	0.827	(0.000)	0.275	(0.000)	R6	7
N005002	0.842	(0.000)	0.859	(0.000)	0.305	(0.000)	R6	8
N005003	0.641	(0.000)	1.372	(0.000)	0.133	(0.000)	R6	9
R000901	1.333	(0.000)	-1.889	(0.000)	0.177	(0.000)	R3	1
R000902	1.298	(0.000)	-1.932	(0.000)	0.177	(0.000)	R3	2
R000903	1.778	(0.000)	-1.758	(0.000)	0.139	(0.000)	R3	3
R000904	1.360	(0.000)	-0.975	(0.000)	0.202	(0.000)	R3	4
R000905	1.526	(0.000)	-1.783	(0.000)	0.139	(0.000)	R3	5
R001501	0.615	(0.000)	-2.105	(0.000)	0.177	(0.000)	R5	1
R001502	0.457	(0.000)	-0.478	(0.000)	0.137	(0.000)	R5	2
R001503	1.066	(0.000)	-1.890	(0.000)	0.197	(0.000)	R5	3
R001504	1.022	(0.000)	-1.054	(0.000)	0.236	(0.000)	R5	4
R001601	1.446	(0.000)	-1.575	(0.000)	0.150	(0.000)	R5	5
R001602	1.173	(0.000)	-1.324	(0.000)	0.203	(0.000)	R5	6
R001603	0.966	(0.000)	-1.142	(0.000)	0.157	(0.000)	R5	7
R001604	0.616	(0.000)	-1.881	(0.000)	0.157	(0.000)	R5	8
R001605	0.728	(0.000)	-0.234	(0.000)	0.189	(0.000)	R5	9
R001701	0.848	(0.000)	-0.831	(0.000)	0.196	(0.000)	R5	10
R001702	1.688	(0.000)	-1.389	(0.000)	0.245	(0.000)	R5	11
R001703	0.587	(0.000)	0.401	(0.000)	0.203	(0.000)	R5	12
R001704	0.724	(0.000)	-0.263	(0.000)	0.166	(0.000)	R5	13
R001801	1.143	(0.000)	-0.850	(0.000)	0.187	(0.000)	R5	14
R001802	1.649	(0.000)	-0.498	(0.000)	0.104	(0.000)	R5	15
R001803	0.637	(0.000)	-0.015	(0.000)	0.202	(0.000)	R5	16
R001804	1.966	(0.000)	-0.448	(0.000)	0.279	(0.000)	R5	17
R001805	1.354	(0.000)	-0.438	(0.000)	0.225	(0.000)	R5	18
R001806	0.654	(0.000)	-0.750	(0.000)	0.146	(0.000)	R5	19
R002501	0.967	(0.000)	-0.292	(0.000)	0.198	(0.000)	R3	6
R002502	0.790	(0.000)	-0.496	(0.000)	0.252	(0.000)	R3	7
R002503	1.049	(0.000)	-1.175	(0.000)	0.131	(0.000)	R3	8
R002601	1.124	(0.000)	-0.114	(0.000)	0.245	(0.000)	R3	9
R002602	1.388	(0.000)	-0.646	(0.000)	0.305	(0.000)	R3	10
R002603	1.367	(0.000)	-1.022	(0.000)	0.263	(0.000)	R3	11
R002604	0.952	(0.000)	-0.150	(0.000)	0.160	(0.000)	R3	12
R002605	0.945	(0.000)	0.049	(0.000)	0.209	(0.000)	R3	13
R002701	1.055	(0.000)	-0.606	(0.000)	0.121	(0.000)	R3	14
R002702	0.777	(0.000)	0.610	(0.000)	0.151	(0.000)	R3	15
R002703	1.487	(0.000)	-0.387	(0.000)	0.181	(0.000)	R3	16
R002704	0.642	(0.000)	0.169	(0.000)	0.173	(0.000)	R3	17
R002705	1.004	(0.000)	1.321	(0.000)	0.214	(0.000)	R3	18
R010001	0.990	(0.036)	-0.727	(0.053)	0.223	(0.025)	R4	1
R010002	0.945	(0.038)	-1.023	(0.069)	0.230	(0.034)	R4	2
R010003	1.391	(0.052)	-0.839	(0.040)	0.267	(0.023)	R4	3
R010004	1.927	(0.095)	-1.371	(0.041)	0.258	(0.030)	R4	4

Table E-14 (continued)
1990 IRT Parameters, Reading Cross-sectional Sample, Age 17/Grade 12

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 17/Grade 12 Block	Item
R010005	0.962	(0.031)	-0.861	(0.049)	0.128	(0.026)	R4	5
R010006	1.039	(0.038)	-0.815	(0.051)	0.194	(0.027)	R4	6
R010007	0.560	(0.014)	-0.634	(0.026)	0.000	(0.000)	R4	7
R010101	1.125	(0.043)	0.304	(0.025)	0.133	(0.012)	R4	8
R010102	1.116	(0.050)	-0.140	(0.043)	0.321	(0.018)	R4	9
R010103	1.337	(0.058)	0.348	(0.023)	0.190	(0.011)	R4	10
R010104	1.295	(0.051)	0.060	(0.028)	0.213	(0.013)	R4	11
R010105	0.838	(0.038)	0.015	(0.052)	0.225	(0.020)	R4	12
R010201	0.745	(0.041)	-0.620	(0.098)	0.294	(0.035)	R4	13
R010202	0.475	(0.046)	0.937	(0.113)	0.261	(0.030)	R4	14
R010203	0.819	(0.040)	-0.493	(0.074)	0.253	(0.029)	R4	15
R010601	0.957	(0.043)	-2.067	(0.102)	0.248	(0.055)	R7	1
R010602	0.812	(0.032)	-1.932	(0.104)	0.234	(0.052)	R7	2
R010603	1.114	(0.043)	-1.197	(0.058)	0.218	(0.032)	R7	3
R010801	1.064	(0.040)	-0.825	(0.050)	0.200	(0.026)	R7	4
R010802	0.953	(0.044)	-0.190	(0.051)	0.310	(0.020)	R7	5
R010803	0.744	(0.034)	-0.929	(0.095)	0.249	(0.037)	R7	6
R010804	0.709	(0.035)	-0.375	(0.081)	0.269	(0.028)	R7	7
R010805	0.437	(0.013)	-0.673	(0.031)	0.000	(0.000)	R7	8
R010901	1.130	(0.036)	-0.338	(0.030)	0.113	(0.015)	R7	9
R010902	0.778	(0.050)	0.796	(0.045)	0.274	(0.015)	R7	10
R010903	0.476	(0.023)	-0.562	(0.123)	0.163	(0.036)	R7	11
R010904	1.015	(0.036)	-0.644	(0.045)	0.155	(0.022)	R7	12
R010905	0.855	(0.039)	0.135	(0.043)	0.169	(0.018)	R7	13
R010906	1.297	(0.049)	-0.171	(0.028)	0.150	(0.014)	R7	14
R011201	0.914	(0.026)	-1.649	(0.038)	0.250	(0.000)	R8	9
R011202	1.228	(0.050)	-1.132	(0.054)	0.249	(0.030)	R8	10
R011203	0.524	(0.045)	0.960	(0.083)	0.246	(0.025)	R8	11
R011204	0.560	(0.037)	0.521	(0.080)	0.210	(0.025)	R8	12
R011301	0.962	(0.034)	-0.085	(0.036)	0.146	(0.016)	R8	13
R011302	1.425	(0.054)	-0.607	(0.033)	0.199	(0.018)	R8	14
R011303	0.935	(0.062)	0.724	(0.041)	0.340	(0.014)	R8	15
R011304	1.733	(0.076)	-0.332	(0.026)	0.281	(0.014)	R8	16
R011305	1.433	(0.053)	0.017	(0.024)	0.199	(0.012)	R8	17
R011401	1.162	(0.067)	0.689	(0.031)	0.358	(0.011)	R8	1
R011402	1.389	(0.065)	-0.639	(0.044)	0.383	(0.021)	R8	2
R011403	1.162	(0.042)	0.074	(0.026)	0.149	(0.013)	R8	3
R011404	0.802	(0.037)	0.041	(0.054)	0.260	(0.020)	R8	4
R011405	0.408	(0.036)	0.913	(0.125)	0.207	(0.032)	R8	5
R011406	1.966	(0.094)	0.497	(0.017)	0.264	(0.008)	R8	6
R011407	1.035	(0.043)	0.302	(0.028)	0.164	(0.013)	R8	7
R011408	1.158	(0.054)	0.682	(0.026)	0.244	(0.010)	R8	8
R011602	0.602	(0.026)	-0.313	(0.076)	0.127	(0.027)	R9	2
R011603	0.664	(0.026)	-1.217	(0.099)	0.188	(0.040)	R9	3
R011604	0.872	(0.036)	-0.433	(0.058)	0.235	(0.024)	R9	4
R011605	1.330	(0.050)	-0.914	(0.042)	0.218	(0.024)	R9	5
R011606	1.532	(0.069)	0.325	(0.023)	0.317	(0.010)	R9	6
R011607	0.892	(0.035)	-0.885	(0.067)	0.216	(0.031)	R9	7
R011701	0.805	(0.031)	-0.792	(0.068)	0.182	(0.030)	R9	8
R011702	0.903	(0.036)	-1.353	(0.082)	0.219	(0.041)	R9	9
R011703	0.626	(0.028)	-0.336	(0.081)	0.163	(0.029)	R9	10
R011801	0.792	(0.039)	0.074	(0.055)	0.245	(0.020)	R9	11
R011802	1.759	(0.073)	0.293	(0.019)	0.264	(0.009)	R9	12
R011803	1.563	(0.066)	-0.232	(0.028)	0.321	(0.014)	R9	13
R011804	1.269	(0.052)	0.211	(0.025)	0.200	(0.012)	R9	14
R011805	0.657	(0.020)	1.149	(0.029)	0.000	(0.000)	R9	15

Table E-15
1990 IRT Parameters, Mathematics Trend Sample, Age 9

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 9 Block	Age 9 Item
N250301	1.517	(0.277)	1.164	(0.060)	0.341	(0.017)	M2	20
N250601	0.920	(0.068)	-0.815	(0.100)	0.196	(0.045)	M2	13
N250602	0.762	(0.052)	-0.837	(0.112)	0.175	(0.045)	M2	14
N250603	1.004	(0.085)	0.297	(0.058)	0.129	(0.027)	M2	15
N250701	0.740	(0.055)	-1.186	(0.128)	0.166	(0.050)	M1	7
N250702	1.311	(0.156)	0.594	(0.044)	0.163	(0.021)	M1	8
N250703	1.065	(0.078)	-0.392	(0.071)	0.122	(0.034)	M1	9
N250901	0.646	(0.052)	-1.083	(0.159)	0.219	(0.055)	M2	17
N250902	1.269	(0.141)	0.741	(0.041)	0.165	(0.019)	M2	18
N250903	1.166	(0.093)	0.043	(0.053)	0.118	(0.027)	M2	19
N251401	0.712	(0.057)	-0.520	(0.124)	0.189	(0.047)	M2	16
N252001	2.038	(0.540)	1.738	(0.096)	0.215	(0.011)	M2	25
N252101	0.671	(0.158)	2.296	(0.254)	0.195	(0.023)	M1	25
N257201	1.081	(0.088)	-0.555	(0.091)	0.240	(0.041)	M1	11
N257801	0.669	(0.058)	-0.990	(0.174)	0.303	(0.057)	M2	3
N258501	0.544	(0.170)	1.856	(0.303)	0.259	(0.040)	M3	19
N261401	0.524	(0.048)	-0.330	(0.164)	0.212	(0.048)	M2	12
N262201	0.717	(0.071)	-0.513	(0.158)	0.308	(0.051)	M1	10
N262401	0.698	(0.153)	0.347	(0.165)	0.282	(0.055)	M3	18
N262501	0.757	(0.148)	0.775	(0.135)	0.490	(0.034)	M1	19
N263401	0.929	(0.081)	-0.975	(0.132)	0.303	(0.054)	M2	4
N263402	0.931	(0.082)	-0.481	(0.117)	0.290	(0.048)	M2	5
N265401	0.755	(0.250)	3.082	(0.538)	0.318	(0.017)	M1	21
N266101	0.720	(0.177)	1.967	(0.207)	0.299	(0.024)	M1	22
N267001	0.759	(0.093)	-1.521	(0.192)	0.278	(0.063)	M3	16
N267601	1.258	(0.116)	-0.507	(0.089)	0.323	(0.039)	M1	3
N267602	1.043	(0.098)	0.087	(0.070)	0.207	(0.030)	M1	18
N268201	1.024	(0.110)	0.575	(0.055)	0.162	(0.024)	M1	24
N269001	0.780	(0.160)	2.881	(0.274)	0.093	(0.012)	M2	26
N269101	0.495	(0.109)	2.061	(0.216)	0.210	(0.033)	M1	23
N270001	0.605	(0.032)	-0.592	(0.046)	0.000	(0.000)	M1	14
N270901	0.861	(0.063)	-2.535	(0.119)	0.000	(0.000)	M1	1
N271101	0.729	(0.033)	-0.965	(0.031)	0.000	(0.000)	M2	24
N272101	0.980	(0.161)	-0.712	(0.162)	0.287	(0.061)	M3	17
N272102	0.877	(0.084)	-0.138	(0.099)	0.197	(0.040)	M1	15
N272301	0.881	(0.069)	-2.189	(0.139)	0.180	(0.053)	M2	1
N272801	0.758	(0.089)	-1.760	(0.177)	0.195	(0.057)	M3	15
N273501	0.696	(0.088)	-0.448	(0.209)	0.457	(0.053)	M2	6
N275401	0.940	(0.040)	-0.770	(0.035)	0.000	(0.000)	M2	7
N276001	1.083	(0.047)	-0.798	(0.030)	0.000	(0.000)	M2	21
N276002	0.899	(0.047)	1.043	(0.042)	0.000	(0.000)	M2	22
N276101	1.046	(0.049)	-0.815	(0.035)	0.000	(0.000)	M1	12
N276601	1.124	(0.098)	-1.040	(0.109)	0.283	(0.051)	M2	2
N276801	0.399	(0.051)	-5.026	(0.578)	0.000	(0.000)	M1	4
N276802	0.645	(0.039)	-2.231	(0.109)	0.000	(0.000)	M1	5
N276803	0.574	(0.032)	-0.092	(0.040)	0.000	(0.000)	M1	6
N277401	0.657	(0.051)	-2.212	(0.174)	0.190	(0.055)	M1	2
N277501	0.902	(0.037)	-0.433	(0.031)	0.000	(0.000)	M2	8
N277601	0.915	(0.040)	-0.992	(0.041)	0.000	(0.000)	M2	9
N277602	0.751	(0.035)	-0.117	(0.031)	0.000	(0.000)	M2	10
N277603	0.748	(0.035)	-0.333	(0.034)	0.000	(0.000)	M2	11
N284001	0.685	(0.035)	-0.896	(0.048)	0.000	(0.000)	M1	16
N284002	0.728	(0.057)	1.869	(0.108)	0.000	(0.000)	M1	17
N286101	0.763	(0.037)	-0.864	(0.044)	0.000	(0.000)	M1	13
N286102	0.896	(0.039)	0.191	(0.027)	0.000	(0.000)	M2	23

Table E-16
1990 IRT Parameters, Mathematics Trend Sample, Age 13

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 13 Block	Age 13 Item
N250201	0.592	(0.034)	-1.400	(0.149)	0.234	(0.052)	M2	19
N250701	0.548	(0.038)	-3.465	(0.217)	0.108	(0.043)	M2	14
N250702	0.970	(0.038)	-1.259	(0.060)	0.093	(0.034)	M2	15
N250703	0.502	(0.033)	-3.306	(0.217)	0.138	(0.048)	M2	16
N250901	0.466	(0.030)	-2.602	(0.196)	0.160	(0.051)	M1	25
N250902	0.807	(0.039)	-0.726	(0.072)	0.116	(0.033)	M1	26
N250903	0.935	(0.043)	-1.858	(0.080)	0.110	(0.039)	M1	27
N252001	1.005	(0.078)	0.797	(0.043)	0.204	(0.017)	M2	40
N252101	0.761	(0.063)	0.483	(0.080)	0.252	(0.028)	M1	41
N252901	0.968	(0.049)	0.073	(0.047)	0.120	(0.021)	M1	32
N253701	0.371	(0.031)	-0.753	(0.251)	0.235	(0.056)	M2	22
N254001	0.790	(0.056)	-0.794	(0.107)	0.153	(0.047)	M3	28
N254601	0.830	(0.061)	-2.219	(0.147)	0.251	(0.058)	M1	16
N254602	0.705	(0.065)	1.097	(0.067)	0.192	(0.022)	M1	46
N255701	1.271	(0.099)	1.177	(0.035)	0.147	(0.012)	M1	50
N256101	0.790	(0.028)	-1.322	(0.038)	0.000	(0.000)	M2	17
N256501	0.998	(0.134)	0.413	(0.085)	0.226	(0.037)	M3	30
N256801	1.326	(0.227)	0.656	(0.056)	0.279	(0.025)	M3	32
N257601	1.131	(0.037)	-0.771	(0.022)	0.000	(0.000)	M1	35
N258801	0.966	(0.103)	1.003	(0.061)	0.355	(0.019)	M1	38
N258802	1.547	(0.112)	0.587	(0.029)	0.202	(0.014)	M2	31
N258803	1.593	(0.118)	1.145	(0.027)	0.150	(0.008)	M2	41
N260101	1.649	(0.112)	0.244	(0.032)	0.220	(0.017)	M1	43
N261001	0.814	(0.059)	0.412	(0.069)	0.229	(0.025)	M1	47
N261201	0.412	(0.070)	2.200	(0.159)	0.204	(0.032)	M2	38
N261301	0.422	(0.038)	1.149	(0.110)	0.110	(0.029)	M2	37
N261501	0.625	(0.032)	-1.065	(0.112)	0.149	(0.043)	M2	34
N261801	0.551	(0.036)	-0.306	(0.128)	0.180	(0.041)	M2	35
N262201	0.415	(0.030)	-2.191	(0.245)	0.256	(0.060)	M2	18
N262401	0.995	(0.056)	-0.602	(0.072)	0.219	(0.034)	M1	28
N262501	0.411	(0.033)	-1.082	(0.248)	0.275	(0.059)	M1	33
N263101	0.583	(0.023)	-0.530	(0.032)	0.000	(0.000)	M1	39
N263401	0.749	(0.045)	-2.406	(0.142)	0.216	(0.056)	M2	12
N263402	0.626	(0.036)	-2.154	(0.152)	0.215	(0.055)	M2	13
N263501	1.009	(0.046)	0.019	(0.043)	0.075	(0.021)	M2	30
N26[illegible]01	1.381	(0.087)	0.483	(0.032)	0.168	(0.016)	M2	33
N265201	0.713	(0.049)	-2.222	(0.162)	0.256	(0.058)	M1	36
N265202	0.600	(0.041)	-0.471	(0.134)	0.222	(0.045)	M1	30
N265901	0.834	(0.075)	0.850	(0.066)	0.281	(0.022)	M1	40
N265902	0.678	(0.095)	0.695	(0.119)	0.262	(0.040)	M3	31
N266101	0.837	(0.076)	-0.676	(0.130)	0.234	(0.054)	M3	27
N266801	0.593	(0.035)	-1.435	(0.151)	0.219	(0.054)	M1	31
N267201	0.735	(0.050)	-1.213	(0.134)	0.265	(0.052)	M1	23
N269001	1.022	(0.068)	0.285	(0.049)	0.188	(0.022)	M1	44
N269101	0.904	(0.059)	-0.245	(0.071)	0.178	(0.032)	M2	26
N269201	1.215	(0.047)	1.508	(0.035)	0.000	(0.000)	M2	44
N269901	0.571	(0.048)	-0.604	(0.168)	0.217	(0.054)	M3	29
N270301	0.368	(0.026)	-2.539	(0.220)	0.119	(0.047)	M2	20
N270302	1.177	(0.104)	1.879	(0.056)	0.049	(0.006)	M2	21
N273901	1.685	(0.104)	0.136	(0.030)	0.216	(0.015)	M1	37
N274801	0.840	(0.075)	0.136	(0.088)	0.343	(0.030)	M1	29
N275001	0.884	(0.028)	0.779	(0.026)	0.000	(0.000)	M1	42
N275301	0.344	(0.030)	-2.886	(0.291)	0.161	(0.053)	M3	25
N276801	0.537	(0.039)	-3.904	(0.241)	0.000	(0.000)	M1	17

Table E-16 (continued)
1990 IRT Parameters, Mathematics Trend Sample, Age 13

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 13 Block	Age 13 Item
N276802	0.455	(0.039)	-4.319	(0.319)	0.000	(0.000)	M1	18
N276803	0.423	(0.021)	-2.098	(0.102)	0.000	(0.000)	M1	19
N277401	0.653	(0.044)	-3.388	(0.204)	0.152	(0.050)	M2	8
N277601	0.552	(0.032)	-3.175	(0.159)	0.000	(0.000)	M1	20
N277602	0.506	(0.023)	-2.100	(0.089)	0.000	(0.000)	M1	21
N277603	0.642	(0.029)	-2.089	(0.080)	0.000	(0.000)	M1	22
N277901	0.741	(0.037)	-2.881	(0.112)	0.000	(0.000)	M2	9
N277902	0.783	(0.046)	-3.324	(0.145)	0.000	(0.000)	M2	10
N277903	0.667	(0.033)	-2.627	(0.102)	0.000	(0.000)	M2	11
N278901	1.363	(0.080)	0.300	(0.033)	0.191	(0.016)	M2	32
N278902	1.108	(0.108)	1.058	(0.047)	0.270	(0.016)	M2	29
N278903	1.882	(0.145)	0.877	(0.026)	0.192	(0.011)	M2	42
N278904	1.014	(0.132)	1.835	(0.069)	0.205	(0.012)	M1	49
N281401	0.908	(0.100)	1.893	(0.072)	0.153	(0.011)	M2	39
N281901	0.903	(0.063)	-2.521	(0.136)	0.169	(0.052)	M1	15
N282201	1.372	(0.100)	0.533	(0.035)	0.270	(0.015)	M2	28
N282202	0.687	(0.062)	-0.631	(0.152)	0.232	(0.056)	M3	26
N283101	2.181	(0.204)	1.282	(0.024)	0.127	(0.006)	M1	51
N285701	0.512	(0.042)	-0.156	(0.155)	0.184	(0.046)	M2	27
N286201	0.868	(0.043)	-0.893	(0.082)	0.169	(0.040)	M1	24
N286301	1.090	(0.061)	0.490	(0.041)	0.159	(0.017)	M1	45
N286501	1.117	(0.081)	1.030	(0.036)	0.131	(0.013)	M1	48
N286502	1.184	(0.078)	1.180	(0.031)	0.085	(0.010)	M2	43
N286601	1.256	(0.036)	-0.379	(0.017)	0.000	(0.000)	M2	23
N286602	1.069	(0.031)	-0.485	(0.020)	0.000	(0.000)	M2	24
N286603	1.116	(0.033)	0.496	(0.018)	0.000	(0.000)	M2	25

Table E-17
1990 IRT Parameters, Mathematics Trend Sample, Age 17

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 17 Block	Item
N251101	1.394	(0.046)	1.021	(0.023)	0.000	(0.000)	M1	49
N251701	1.034	(0.065)	-0.016	(0.059)	0.222	(0.026)	M2	41
N253901	1.259	(0.073)	-0.260	(0.052)	0.247	(0.026)	M1	39
N253902	0.602	(0.077)	0.730	(0.149)	0.406	(0.036)	M1	40
N253903	1.075	(0.080)	0.603	(0.052)	0.280	(0.019)	M1	41
N253904	1.561	(0.137)	0.640	(0.039)	0.363	(0.015)	M1	42
N254001	0.890	(0.056)	-0.898	(0.103)	0.259	(0.047)	M2	21
N254301	1.162	(0.093)	0 299	(0.055)	0.350	(0.021)	M1	33
N254601	0.998	(0.059)	-2.511	(0.116)	0.208	(0.058)	M2	15
N254602	1.291	(0.068)	-0.343	(0.045)	0.159	(0.026)	M1	27
N255501	0.668	(0.084)	0.103	(0.163)	0.224	(0.055)	M3	33
N255601	1.766	(0.261)	1.598	(0.048)	0.340	(0.011)	M2	45
N255701	1.348	(0.066)	-0.780	(0.047)	0.145	(0.030)	M1	32
N255801	0.847	(0.032)	1.600	(0.047)	0.000	(0.000)	M2	49
N256001	0.636	(0.037)	-0.178	(0.030)	0.000	(0.000)	M3	34
N256101	0.920	(0.041)	-1.969	(0.060)	0.000	(0.000)	M1	15
N256801	1.062	(0.054)	-0.522	(0.062)	0.194	(0.032)	M1	36
N257101	0.468	(0.087)	1.450	(0.167)	0.190	(0.045)	M3	35
N258801	1.032	(0.066)	-0.539	(0.079)	0.250	(0.038)	M2	38
N258802	1.580	(0.098)	-0.489	(0.044)	0.209	(0.028)	M1	26
N258803	0.973	(0.068)	0.050	(0.062)	0.221	(0.027)	M1	37
N258804	0.671	(0.052)	-2.628	(0.182)	0.212	(0.058)	M1	18
N259001	0.948	(0.031)	-0.443	(0.022)	0.000	(0.000)	M2	31
N259901	0.927	(0.051)	-0.555	(0.074)	0.177	(0.036)	M1	28
N260101	1.243	(0.058)	-1.213	(0.061)	0.165	(0.041)	M2	20
N260601	1.821	(0.089)	-1.420	(0.025)	0.000	(0.000)	M1	16
N260801	1.210	(0.034)	0.114	(0.018)	0.000	(0.000)	M2	43
N260901	1.672	(0.103)	-0.048	(0.033)	0.168	(0.020)	M1	35
N261001	0.735	(0.045)	-0.700	(0.111)	0.209	(0.045)	M2	40
N261201	0.517	(0.039)	-0.212	(0.157)	0.203	(0.047)	M2	26
N261301	0.679	(0.050)	0.423	(0.083)	0.179	(0.029)	M2	28
N261501	0.676	(0.042)	-2.036	(0.148)	0.232	(0.060)	M2	24
N261601	0.666	(0.101)	1.893	(0.110)	0.343	(0.019)	M2	27
N261801	0.603	(0.035)	-1.451	(0.141)	0.196	(0.053)	M2	25
N262301	0.546	(0.038)	-1.601	(0.179)	0.225	(0.059)	M2	17
N262401	1.121	(0.058)	-1.354	(0.080)	0.212	(0.048)	M1	17
N262501	0.552	(0.036)	-1.218	(0.194)	0.321	(0.059)	M2	35
N262502	1.117	(0.122)	1.428	(0.054)	0.317	(0.012)	M2	36
N262601	0.608	(0.047)	0.356	(0.104)	0.180	(0.034)	M1	38
N263001	0.517	(0.022)	0.922	(0.045)	0.000	(0.000)	M1	43
N263101	0.662	(0.025)	-0.867	(0.035)	0.000	(0.000)	M2	37
N263201	0.793	(0.053)	-1.622	(0.154)	0.356	(0.062)	M2	18
N263202	0.725	(0.055)	-0.771	(0.148)	0.348	(0.051)	M2	19
N264301	0.819	(0.029)	0.988	(0.032)	0.000	(0.000)	M1	47
N264701	1.203	(0.060)	-0.396	(0.051)	0.158	(0.030)	M2	39
N266501	0.733	(0.071)	-0.461	(0.149)	0.219	(0.057)	M3	31
N268801	1.151	(0.092)	1.440	(0.040)	0.113	(0.009)	M2	48
N268901	1.532	(0.094)	0.541	(0.030)	0.183	(0.013)	M2	47
N269001	1.321	(0.089)	-0.048	(0.046)	0.274	(0.022)	M2	22
N270301	0.854	(0.052)	-1.681	(0.112)	0.192	(0.056)	M1	30
N270302	1.342	(0.058)	0.042	(0.030)	0.095	(0.015)	M1	31
N271301	1.418	(0.265)	-0.010	(0.097)	0.233	(0.048)	M3	32
N278501	0.787	(0.031)	-1.155	(0.038)	0.000	(0.000)	M1	23
N278502	0.825	(0.030)	-0.883	(0.030)	0.000	(0.000)	M1	24

Table E-17 (continued)
1990 IRT Parameters, Mathematics Trend Sample, Age 17

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 17 Block	Age 17 Item
N278503	0.635	(0.028)	-1.406	(0.053)	0.000	(0.000)	M1	25
N278901	1.013	(0.049)	-0.766	(0.065)	0.157	(0.035)	M2	23
N278902	0.845	(0.049)	-0.442	(0.083)	0.187	(0.037)	M2	42
N278903	1.206	(0.068)	-0.078	(0.049)	0.208	(0.025)	M2	44
N278905	0.638	(0.067)	1.190	(0.086)	0.194	(0.027)	M1	44
N280401	0.446	(0.022)	-1.486	(0.073)	0.000	(0.000)	M2	30
N281401	0.578	(0.059)	1.311	(0.086)	0.170	(0.025)	M2	29
N286001	0.765	(0.043)	-1.014	(0.106)	0.197	(0.047)	M1	19
N286002	1.097	(0.054)	-1.557	(0.074)	0.148	(0.045)	M1	20
N286301	1.223	(0.059)	-0.666	(0.055)	0.188	(0.032)	M2	33
N286302	1.214	(0.078)	-0.551	(0.073)	0.292	(0.038)	M1	22
N286501	1.374	(0.077)	-0.755	(0.057)	0.244	(0.034)	M2	34
N286502	1.587	(0.084)	-0.255	(0.035)	0.177	(0.021)	M1	34
N287101	1.168	(0.074)	-0.380	(0.059)	0.232	(0.030)	M1	29
N287102	1.159	(0.057)	-0.612	(0.054)	0.168	(0.030)	M2	32
N287301	0.629	(0.023)	-0.115	(0.029)	0.000	(0.000)	M1	45
N287302	0.777	(0.029)	1.342	(0.043)	0.000	(0.000)	M1	46

Table E-18
1990 IRT Parameters, Mathematics Cross-sectional Samples
Numbers and Operations Subscale

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 9/Grade 4 Block	Age 9/Grade 4 Item	Age 13/Grade 8 Block	Age 13/Grade 8 Item	Age 17/Grade 12 Block	Age 17/Grade 12 Item
M010131	0.753	(0.049)	-1.783	(0.122)	0.210	(0.054)	M8	2	--	--	--	--
M010231	0.665	(0.048)	-1.500	(0.131)	0.200	(0.051)	M8	3	--	--	--	--
M010431	0.771	(0.071)	-0.613	(0.086)	0.199	(0.036)	M8	5	--	--	--	--
M010531	0.805	(0.139)	0.667	(0.105)	0.209	(0.021)	M8	6	--	--	--	--
M010631	0.541	(0.028)	-1.919	(0.055)	0.000	(0.000)	M8	7	--	--	--	--
M010831	1.350	(0.091)	-0.526	(0.034)	0.169	(0.019)	M8	9	--	--	--	--
M011131	1.003	(0.040)	-0.669	(0.048)	0.176	(0.024)	M8	13	M8	13	--	--
M011531	4.260	(0.000)	1.687	(0.017)	0.104	(0.000)	--	--	--	--	M8	15
M012431	0.645	(0.040)	-0.526	(0.124)	0.169	(0.052)	--	--	--	--	M8	3
M012431	1.252	(0.086)	0.287	(0.045)	0.153	(0.027)	--	--	M8	3	--	--
M012531	1.101	(0.052)	0.802	(0.029)	0.130	(0.016)	--	--	M8	4	M8	4
M012931	1.503	(0.094)	1.185	(0.021)	0.227	(0.010)	--	--	M8	8	M8	8
M013431	1.808	(0.139)	0.619	(0.028)	0.212	(0.016)	--	--	M8	15	--	--
M013531	1.209	(0.170)	1.549	(0.064)	0.116	(0.013)	--	--	M8	16	--	--
M013631	2.453	(0.229)	1.050	(0.020)	0.099	(0.008)	--	--	M8	17	--	--
M014601	1.718	(0.127)	-0.600	(0.030)	0.204	(0.018)	M7	10	--	--	--	--
M014701	1.240	(0.103)	-0.764	(0.052)	0.217	(0.029)	M7	11	--	--	--	--
M014801	1.445	(0.126)	-0.263	(0.032)	0.219	(0.017)	M7	12	--	--	--	--
M014901	1.501	(0.101)	-0.480	(0.028)	0.138	(0.017)	M7	13	--	--	--	--
M015101	1.802	(0.390)	0.522	(0.079)	0.335	(0.010)	M7	15	--	--	--	--
M015201	1.826	(0.139)	-0.135	(0.024)	0.168	(0.011)	M7	16	--	--	--	--
M015301	0.662	(0.076)	-0.665	(0.137)	0.273	(0.047)	M7	18	--	--	--	--
M015501	1.529	(0.065)	0.569	(0.021)	0.122	(0.014)	--	--	M7	2	M7	2
M015901	1.327	(0.094)	1.103	(0.027)	0.295	(0.013)	--	--	M7	6	M7	6
M016501	1.568	(0.098)	1.599	(0.024)	0.090	(0.006)	--	--	M7	12	M7	12
M017401	0.442	(0.018)	-2.907	(0.175)	0.200	(0.055)	M4	1	M4	1	M4	1
M017701	1.178	(0.034)	-0.400	(0.029)	0.143	(0.015)	M4	4	M4	4	M4	4
M017901	1.885	(0.051)	-0.221	(0.015)	0.113	(0.009)	M4	6	M4	6	M4	6
M018201	1.153	(0.031)	-0.026	(0.023)	0.121	(0.011)	M4	9	M4	9	M4	9
M018401	1.871	(0.059)	-0.196	(0.018)	0.207	(0.010)	M4	11	M4	11	M4	11
M018501	2.389	(0.113)	0.731	(0.012)	0.228	(0.005)	M4	12	M4	12	M4	12
M018601	0.874	(0.043)	1.198	(0.025)	0.150	(0.008)	M4	13	M4	13	M4	13
M020001	1.729	(0.064)	-0.803	(0.012)	0.000	(0.000)	M5	4	--	--	--	--
M020001	0.788	(0.031)	0.167	(0.020)	0.000	(0.000)	--	--	M5	4	--	--
M020101	1.444	(0.030)	0.050	(0.011)	0.000	(0.000)	M5	5	M5	5	--	--
M020501	1.125	(0.019)	0.035	(0.010)	0.000	(0.000)	M5	9	M5	9	M5	9
M021401	1.204	(0.043)	0.622	(0.015)	0.000	(0.000)	--	--		--	M5	1
M021701	2.023	(0.087)	1.216	(0.014)	0.000	(0.000)	--	--		--	M5	15
M021901	1.011	(0.031)	-0.782	(0.048)	0.181	(0.025)	M6	1	M6	1	M6	1
M022001	1.474	(0.056)	-0.202	(0.039)	0.157	(0.029)	--	--	M6	2	M6	2
M022001	2.433	(0.398)	0.369	(0.041)	0.183	(0.007)	M6	2	--	--	--	--
M022301	1.104	(0.045)	-1.003	(0.054)	0.179	(0.031)	M6	5	M6	5	--	--
M022701	1.871	(0.093)	0.020	(0.022)	0.309	(0.010)	M6	9	M6	9	--	--
M022901	1.555	(0.076)	0.053	(0.024)	0.257	(0.011)	M6	12	M6	12	--	--
M023001	1.520	(0.072)	0.233	(0.022)	0.231	(0.010)	M6	13	M6	13	--	--
M023801	1.685	(0.099)	0.583	(0.022)	0.078	(0.014)	--	--	M6	21	--	--
M023901	0.589	(0.030)	0.268	(0.034)	0.000	(0.000)	--	--	--	--	M6	3
M025101	1.957	(0.406)	2.171	(0.075)	0.252	(0.009)	--	--	--	--	M6	15
M025302	1.250	(0.053)	1.646	(0.030)	0.000	(0.000)	--	--	--	--	M6	17
M025531	0.445	(0.029)	-2.897	(0.121)	0.000	(0.000)	M9	1	--	--	--	--
M025631	1.049	(0.081)	-1.227	(0.085)	0.242	(0.043)	M9	2	--	--	--	--
M025931	0.819	(0.072)	-0.838	(0.092)	0.211	(0.040)	M9	5	--	--	--	--
M026(31	1.841	(0.134)	-0.537	(0.025)	0.192	(0.016)	M9	6	--	--	--	--

Table E-18 (continued)
1990 IRT Parameters, Mathematics Cross-sectional Samples
Numbers and Operations Subscale

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 9/Grade 4 Block	Age 9/Grade 4 Item	Age 13/Grade 8 Block	Age 13/Grade 8 Item	Age 17/Grade 12 Block	Age 17/Grade 12 Item
M026131	1.644	(0.107)	-0.528	(0.027)	0.164	(0.016)	M9	7	–	–	–	–
M026231	0.920	(0.105)	-0.063	(0.057)	0.230	(0.025)	M9	8	–	–	–	–
M026331	1.113	(0.101)	-0.644	(0.060)	0.257	(0.029)	M9	9	–	–	–	–
M026531	1.017	(0.406)	2.110	(0.719)	0.185	(0.008)	M9	11	–	–	–	–
M026731	2.019	(0.247)	0.234	(0.033)	0.131	(0.009)	M9	13	–	–	–	–
M027031	0.861	(0.057)	-1.746	(0.125)	0.205	(0.057)	–	–	M9	1	–	–
M027331	1.130	(0.039)	0.764	(0.019)	0.000	(0.000)	–	–	M9	4	–	–
M027831	1.407	(0.045)	0.373	(0.014)	0.000	(0.000)	–	–	M9	9	–	–
M028031	1.532	(0.112)	0.714	(0.028)	0.176	(0.016)	–	–	M9	11	–	–
M028131	0.822	(0.035)	0.879	(0.028)	0.000	(0.000)	–	–	M9	12	–	–
M028231	1.064	(0.072)	0.675	(0.038)	0.090	(0.020)	–	–	M9	13	–	–
M028631	1.975	(0.103)	1.271	(0.024)	0.000	(0.000)	–	–	M9	17	–	–
M028731	2.837	(0.368)	1.305	(0.025)	0.088	(0.006)	–	–	M9	18	–	–
M028931	0.583	(0.079)	1.038	(0.118)	0.167	(0.042)	–	–	M9	20	–	–
M029031	2.330	(0.161)	0.494	(0.031)	0.244	(0.024)	–	–	–	–	M9	1
M029131	1.282	(0.063)	-0.198	(0.063)	0.164	(0.047)	–	–	–	–	M9	2
M029231	2.849	(0.231)	1.426	(0.016)	0.107	(0.007)	–	–	–	–	M9	3
M029331	1.054	(0.054)	0.140	(0.065)	0.140	(0.040)	–	–	–	–	M9	4
M029931	2.140	(0.182)	1.193	(0.021)	0.215	(0.013)	–	–	–	–	M9	10
M030331	1.207	(0.134)	1.686	(0.042)	0.213	(0.017)	–	–	–	–	M9	13
M030531	2.421	(0.355)	2.020	(0.043)	0.099	(0.006)	–	–	–	–	M9	15
N202831	1.268	(0.080)	-0.347	(0.080)	0.221	(0.058)	–	–	–	–	M8	12
N202831	0.925	(0.047)	-1.132	(0.083)	0.265	(0.039)	M8	12	M8	12	–	–
N230501	1.580	(0.123)	-0.490	(0.033)	0.232	(0.018)	M7	17	–	–	–	–
N240031	1.965	(0.135)	-0.416	(0.022)	0.153	(0.014)	M8	14	–	–	–	–
N257201	1.093	(0.068)	-1.145	(0.060)	0.167	(0.034)	M3	5	–	–	–	–
N257801	0.788	(0.049)	-1.771	(0.106)	0.175	(0.049)	M3	8	–	–	–	–
N258501	1.013	(0.106)	-0.115	(0.046)	0.141	(0.023)	M3	15	–	–	–	–
N258801	1.747	(0.183)	0.802	(0.034)	0.374	(0.016)	–	–	M3	11	–	–
N258802	3.065	(0.255)	0.500	(0.025)	0.238	(0.021)	–	–	–	–	M3	12
N258804	1.216	(0.071)	-0.661	(0.078)	0.182	(0.055)	–	–	–	–	M3	1
N259001	1.301	(0.044)	0.505	(0.015)	0.000	(0.000)	–	–	–	–	M3	14
N260101	1.556	(0.103)	0.171	(0.039)	0.170	(0.025)	–	–	M3	18	–	–
N260601	1.877	(0.086)	-0.280	(0.023)	0.000	(0.000)	–	–	–	–	M3	6
N261401	0.619	(0.047)	-1.136	(0.119)	0.160	(0.045)	M3	19	–	–	–	–
N268201	1.731	(0.116)	-0.488	(0.025)	0.149	(0.015)	M3	13	–	–	–	–
N271301	2.476	(0.177)	0.809	(0.024)	0.258	(0.017)	–	–	–	–	M3	13
N274801	0.734	(0.052)	-0.556	(0.127)	0.205	(0.056)	–	–	M3	10	–	–
N275301	0.403	(0.032)	-1.592	(0.213)	0.165	(0.053)	–	–	M3	14	–	–
N275401	1.086	(0.041)	-1.583	(0.024)	0.000	(0.000)	M3	14	–	–	–	–
N276803	0.406	(0.028)	-1.720	(0.128)	0.000	(0.000)	–	–	M3	1	–	–
N277401	1.032	(0.061)	-2.170	(0.089)	0.165	(0.051)	M3	1	–	–	–	–
N277601	0.951	(0.039)	-1.915	(0.034)	0.000	(0.000)	M3	11	–	–	–	–
N277602	0.779	(0.018)	-1.183	(0.022)	0.000	(0.000)	M3	12	M3	2	–	–
N277903	0.915	(0.038)	-1.630	(0.031)	0.000	(0.000)	M5	10	–	–	–	–
N278501	1.006	(0.038)	-0.013	(0.026)	0.000	(0.000)	–	–	–	–	M3	8
N286201	1.135	(0.069)	-0.355	(0.075)	0.188	(0.045)	–	–	M3	6	–	–
N286301	2.100	(0.158)	0.488	(0.024)	0.212	(0.016)	–	–	M3	21	–	–
N286302	1.546	(0.102)	0.267	(0.059)	0.236	(0.042)	–	–	–	–	M3	10
N286602	0.899	(0.033)	0.299	(0.019)	0.000	(0.000)	–	–	M3	13	–	–

Table E-19
1990 IRT Parameters, Mathematics Cross-sectional Samples
Measurement Subscale

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 9/Grade 4 Block	Age 9/Grade 4 Item	Age 13/Grade 8 Block	Age 13/Grade 8 Item	Age 17/Grade 12 Block	Age 17/Grade 12 Item
M010731	1.895	(0.407)	0.049	(0.045)	0.307	(0.020)	M8	8	--	--	--	--
M010931	0.933	(0.195)	0.434	(0.089)	0.217	(0.031)	M8	10	--	--	--	--
M011331	2.986	(0.549)	1.254	(0.034)	0.249	(0.018)	--	--	--	--	M8	13
M011931	2.607	(0.501)	2.366	(0.071)	0.000	(0.000)	--	--	--	--	M8	19
M012331	0.932	(0.054)	-0.876	(0.100)	0.203	(0.053)	--	--	M8	2	M8	2
M013331	1.184	(0.108)	-0.764	(0.096)	0.210	(0.054)	--	--	M8	14	--	--
M014001	0.918	(0.079)	-1.413	(0.120)	0.243	(0.059)	M7	3	--	--	--	--
M014201	1.612	(0.190)	-0.977	(0.076)	0.232	(0.048)	M7	5	--	--	--	--
M014301	1.369	(0.162)	-0.655	(0.065)	0.197	(0.038)	M7	6	--	--	--	--
M015401	0.944	(0.061)	0.160	(0.068)	0.159	(0.032)	--	--	M7	1	M7	1
M015701	1.025	(0.063)	-1.439	(0.103)	0.223	(0.059)	--	--	M7	4	M7	4
M016201	1.344	(0.113)	1.036	(0.034)	0.218	(0.015)	--	--	M7	9	M7	9
M017501	0.591	(0.026)	-1.616	(0.133)	0.207	(0.053)	M4	2	M4	2	M4	2
M018101	0.801	(0.041)	0.075	(0.059)	0.132	(0.024)	M4	8	M4	8	M4	8
M019101	1.802	(0.181)	1.610	(0.033)	0.189	(0.009)	--	--	M4	18	M4	18
M019201	2.044	(0.244)	1.685	(0.032)	0.152	(0.008)	--	--	M4	19	M4	19
M020301	1.558	(0.043)	-0.021	(0.012)	0.000	(0.000)	M5	7	M5	7	M5	7
M022601	0.947	(0.110)	0.603	(0.056)	0.316	(0.020)	M6	8	M6	8	--	--
M022801	1.633	(0.060)	-0.470	(0.013)	0.000	(0.000)	M6	10	M6	10	--	--
M022802	1.676	(0.062)	-0.723	(0.014)	0.000	(0.000)	M6	11	M6	11	--	--
M023401	1.081	(0.157)	0.351	(0.088)	0.343	(0.035)	--	--	M6	17	--	--
M023401	1.473	(0.378)	0.795	(0.139)	0.153	(0.013)	M6	17	--	--	--	--
M023701	0.789	(0.046)	1.010	(0.048)	0.000	(0.000)	--	--	M6	20	--	--
M024001	0.735	(0.067)	-0.909	(0.160)	0.225	(0.059)	--	--	--	--	M6	4
M024301	1.464	(0.112)	1.117	(0.038)	0.106	(0.021)	--	--	--	--	M6	7
M024601	0.964	(0.097)	1.110	(0.077)	0.194	(0.035)	--	--	--	--	M6	10
M024701	3.163	(0.262)	1.402	(0.016)	0.000	(0.000)	--	--	--	--	M6	11
M024801	2.165	(0.237)	0.959	(0.047)	0.322	(0.025)	--	--	--	--	M6	12
M026431	1.629	(0.112)	-0.167	(0.025)	0.000	(0.000)	M9	10	--	--	--	--
M026831	1.195	(0.086)	-0.176	(0.033)	0.000	(0.000)	M9	14	--	--	--	--
M027631	1.467	(0.213)	0.309	(0.059)	0.231	(0.030)	--	--	M9	7	--	--
M029731	2.201	(0.531)	2.013	(0.055)	0.215	(0.011)	--	--	--	--	M9	8
M030431	1.680	(0.427)	2.252	(0.091)	0.249	(0.013)	--	--	--	--	M9	14
N215401	1.225	(0.119)	-1.042	(0.090)	0.227	(0.051)	M7	7	--	--	--	--
N252001	2.110	(0.463)	0.351	(0.054)	0.174	(0.015)	M3	17	--	--	--	--
N252101	0.650	(0.087)	0.237	(0.161)	0.251	(0.054)	--	--	M3	17	--	--
N262401	0.932	(0.058)	-0.512	(0.095)	0.258	(0.041)	M3	18	--	--	M3	2
N265201	0.843	(0.076)	-1.613	(0.145)	0.244	(0.061)	--	--	M3	9	--	--
N265901	0.871	(0.148)	0.693	(0.101)	0.289	(0.039)	--	--	M3	16	--	--
N266101	0.918	(0.163)	-0.087	(0.100)	0.276	(0.044)	M3	6	--	--	--	--
N266501	1.424	(0.129)	0.563	(0.069)	0.228	(0.041)	--	--	--	--	M3	11
N267201	0.650	(0.058)	-1.192	(0.166)	0.233	(0.060)	--	--	M3	3	--	--
N267601	1.080	(0.107)	-1.332	(0.116)	0.269	(0.062)	M3	3	--	--	--	--
N268801	2.125	(0.237)	1.410	(0.029)	0.110	(0.015)	--	--	--	--	M3	23
N268901	3.705	(0.657)	1.111	(0.028)	0.170	(0.018)	--	--	--	--	M3	19
N269101	1.312	(0.266)	0.434	(0.076)	0.215	(0.021)	M3	2	--	--	--	--

Table E-20
1990 IRT Parameters, Mathematics Cross-sectional Samples
Geometry Subscale

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 9/Grade 4 Block	Item	Age 13/Grade 8 Block	Item	Age 17/Grade 12 Block	Item
M011231	1.610	(0.804)	0.905	(0.344)	0.272	(0.014)	M8	15	–	–	–	–
M011731	1.679	(0.172)	1.318	(0.038)	0.161	(0.019)	–	–	–	–	M8	17
M012731	1.450	(0.104)	1.060	(0.031)	0.194	(0.014)	–	–	M8	6	M8	6
M012831	1.782	(0.102)	0.552	(0.028)	0.125	(0.016)	–	–	M8	7	M8	7
M014401	1.211	(0.188)	-0.382	(0.069)	0.183	(0.041)	M7	8	–	–	–	–
M014501	1.479	(0.192)	-0.872	(0.072)	0.189	(0.048)	M7	9	–	–	–	–
M015601	0.623	(0.048)	0.081	(0.152)	0.230	(0.053)	–	–	M7	3	M7	3
M016301	0.822	(0.057)	0.227	(0.097)	0.203	(0.042)	–	–	M7	10	M7	10
M016401	1.665	(0.113)	1.144	(0.023)	0.146	(0.010)	–	–	M7	11	M7	11
M016601	1.275	(0.074)	1.157	(0.026)	0.075	(0.011)	–	–	M7	13	M7	13
M016701	2.362	(0.157)	1.165	(0.018)	0.131	(0.007)	–	–	M7	14	M7	14
M017201	1.128	(0.054)	1.237	(0.025)	0.000	(0.000)	–	–	–	–	M7	20
M017601	0.518	(0.029)	-1.119	(0.183)	0.242	(0.059)	M4	3	M4	3	M4	3
M018001	0.873	(0.048)	0.214	(0.057)	0.157	(0.024)	M4	7	M4	7	M4	7
M019001	1.298	(0.084)	0.753	(0.037)	0.164	(0.018)	–	–	M4	17	M4	17
M019601	1.002	(0.188)	1.373	(0.087)	0.157	(0.022)	–	–	M4	21	–	–
M019801	1.281	(0.059)	-0.216	(0.020)	0.000	(0.000)	–	–	M5	2	–	–
M019801	0.709	(0.078)	1.395	(0.209)	0.000	(0.000)	M5	2	–	–	–	–
M019901	0.761	(0.030)	-1.004	(0.027)	0.000	(0.000)	M5	3	M5	3	–	–
M020701	1.229	(0.104)	-0.113	(0.040)	0.000	(0.000)	M5	11	–	–	–	–
M020901	0.574	(0.024)	1.724	(0.051)	0.000	(0.000)	–	–	M5	11	M5	11
M021001	1.188	(0.038)	0.364	(0.015)	0.000	(0.000)	–	–	M5	12	M5	12
M021301	1.741	(0.081)	0.262	(0.016)	0.000	(0.000)	–	–	M5	15	–	–
M021302	1.629	(0.076)	0.152	(0.017)	0.000	(0.000)	–	–	M5	16	–	–
M021801	2.338	(0.183)	1.973	(0.030)	0.000	(0.000)	–	–	–	–	M5	17
M022201	0.827	(0.028)	0.013	(0.020)	0.000	(0.000)	M6	4	M6	4	–	–
M022501	1.429	(0.048)	0.136	(0.015)	0.000	(0.000)	M6	7	M6	7	–	–
M023101	2.419	(0.213)	0.278	(0.021)	0.163	(0.010)	M6	14	M6	14	–	–
M024101	1.441	(0.127)	0.644	(0.065)	0.190	(0.037)	–	–	–	–	M6	5
M024401	1.575	(0.227)	1.697	(0.042)	0.238	(0.016)	–	–	–	–	M6	8
M026931	1.510	(0.155)	-0.569	(0.027)	0.000	(0.000)	M9	15	–	–	–	–
M027231	0.588	(0.059)	-0.730	(0.175)	0.221	(0.061)	–	–	M9	3	–	–
M027431	1.002	(0.089)	-0.115	(0.095)	0.190	(0.049)	–	–	M9	5	–	–
M028331	1.288	(0.296)	1.243	(0.080)	0.316	(0.022)	–	–	M9	14	–	–
M029431	1.262	(0.106)	0.189	(0.086)	0.209	(0.051)	–	–	–	–	M9	5
M029831	2.140	(0.131)	1.685	(0.022)	0.000	(0.000)	–	–	–	–	M9	9
M030131	0.507	(0.117)	3.219	(0.260)	0.189	(0.029)	–	–	–	–	M9	11
N214331	1.378	(0.158)	-1.880	(0.094)	0.206	(0.057)	M8	1	–	–	–	–
N253701	0.564	(0.053)	-0.533	(0.163)	0.196	(0.056)	–	–	M3	12	–	–
N253901	1.641	(0.158)	0.594	(0.062)	0.231	(0.037)	–	–	–	–	M3	15
N253904	1.679	(0.185)	1.204	(0.044)	0.251	(0.022)	–	–	–	–	M3	16
N254001	0.835	(0.063)	0.051	(0.113)	0.188	(0.053)	–	–	–	–	M3	7
N254602	1.674	(0.254)	0.861	(0.035)	0.161	(0.019)	–	–	M3	22	–	–
N255301	1.582	(0.303)	2.274	(0.067)	0.199	(0.012)	–	–	–	–	M3	20
N269901	0.861	(0.078)	-0.196	(0.113)	0.197	(0.053)	–	–	M3	15	–	–
N270001	0.921	(0.074)	-0.895	(0.024)	0.000	(0.000)	M3	7	–	–	–	–
N270301	1.055	(0.076)	-0.618	(0.090)	0.145	(0.050)	–	–	–	–	M3	5

Table E-21
1990 IRT Parameters, Mathematics Cross-sectional Samples
Data Analysis, Statistics, and Probability Subscale

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 9/Grade 4 Block	Age 9/Grade 4 Item	Age 13/Grade 8 Block	Age 13/Grade 8 Item	Age 17/Grade 12 Block	Age 17/Grade 12 Item
M011631	0.895	(0.092)	0.358	(0.084)	0.144	(0.036)	--	--	--	--	M8	16
M012631	1.544	(0.222)	0.184	(0.081)	0.318	(0.037)	--	--	--	--	M8	5
M012631	1.931	(0.372)	0.563	(0.040)	0.210	(0.017)	--	--	M8	5	--	--
M013031	1.296	(0.066)	1.094	(0.029)	0.000	(0.000)	--	--	M8	9	M8	9
M013131	0.910	(0.045)	1.165	(0.039)	0.000	(0.000)	--	--	M8	10	M8	10
M015801	1.409	(0.104)	0.124	(0.040)	0.106	(0.019)	--	--	M7	5	M7	5
M016101	0.997	(0.077)	-0.060	(0.071)	0.160	(0.031)	--	--	M7	8	M7	8
M017001	0.762	(0.117)	0.816	(0.079)	0.147	(0.029)	--	--	M7	18	--	--
M017801	1.105	(0.089)	-0.494	(0.091)	0.225	(0.041)	M4	5	M4	5	M4	5
M018901	1.313	(0.340)	2.381	(0.213)	0.203	(0.008)	--	--	M4	16	M4	16
M020201	0.620	(0.029)	-1.986	(0.076)	0.000	(0.000)	M5	6	M5	6	M5	6
M020801	1.155	(0.065)	1.369	(0.043)	0.000	(0.000)	--	--	M5	10	M5	10
M021101	1.142	(0.044)	-0.036	(0.019)	0.000	(0.000)	--	--	M5	13	M5	13
M021501	1.199	(0.066)	0.061	(0.028)	0.000	(0.000)	--	--	--	--	M5	2
M021502	1.141	(0.064)	-0.568	(0.044)	0.000	(0.000)	--	--	--	--	M5	3
M023301	2.035	(0.316)	-0.606	(0.072)	0.229	(0.038)	M6	16	M6	16	--	--
M023501	1.777	(0.199)	0.663	(0.024)	0.140	(0.012)	--	--	M6	18	M6	18
M023601	0.901	(0.068)	-0.383	(0.091)	0.156	(0.039)	--	--	M6	19	M6	19
M024501	0.412	(0.065)	0.700	(0.236)	0.252	(0.055)	--	--	--	--	M6	9
M028531	1.061	(0.067)	-0.957	(0.041)	0.000	(0.000)	--	--	M9	16	--	--
M029531	1.221	(0.124)	0.203	(0.068)	0.146	(0.033)	--	--	--	--	M9	6
M029631	2.146	(0.546)	1.917	(0.113)	0.051	(0.007)	--	--	--	--	M9	7
M030731	0.970	(0.118)	3.156	(0.245)	0.000	(0.000)	--	--	--	--	M9	17
N250201	0.653	(0.050)	-1.682	(0.139)	0.158	(0.050)	--	--	M3	8	--	--
N250901	0.321	(0.042)	-3.958	(0.494)	0.211	(0.059)	M3	9	M3	4	--	--
N250902	0.847	(0.070)	-0.982	(0.108)	0.148	(0.046)	M3	10	M3	5	--	--
N263001	0.636	(0.044)	0.393	(0.038)	0.000	(0.000)	--	--	--	--	M3	22
N263501	1.294	(0.157)	-0.249	(0.072)	0.182	(0.035)	--	--	M3	19	--	--
N286001	0.670	(0.069)	-0.515	(0.167)	0.221	(0.056)	--	--	--	--	M3	3
N286002	1.047	(0.084)	-1.229	(0.113)	0.148	(0.048)	--	--	--	--	M3	4

Table E-22
1990 IRT Parameters, Mathematics Cross-sectional Samples
Algebra and Functions Subscale

							Age 9/Grade 4		Age 13/Grade 8		Age 17/Grade 12	
NAEP ID	A	S.E.	B	S.E.	C	S.E.	Block	Item	Block	Item	Block	Item
M010331	0.978	(0.172)	-0.483	(0.106)	0.271	(0.050)	M8	4	--	--	--	--
M011431	1.606	(0.147)	1.445	(0.036)	0.206	(0.017)	--	--	--	--	M8	14
M011831	3.956	(0.937)	2.059	(0.033)	0.209	(0.008)	--	--	--	--	M8	18
M012031	1.922	(0.105)	1.511	(0.023)	0.000	(0.000)	--	--	--	--	M8	20
M012131	2.062	(0.292)	2.021	(0.045)	0.157	(0.011)	--	--	--	--	M8	21
M012231	0.640	(0.043)	-2.161	(0.172)	0.193	(0.056)	--	--	M8	1	M8	1
M013231	1.777	(0.149)	1.605	(0.026)	0.155	(0.008)	--	--	M8	11	M8	11
M013731	1.982	(0.331)	1.112	(0.044)	0.157	(0.014)	--	--	M8	18	--	--
M013801	1.260	(0.132)	-1.721	(0.104)	0.253	(0.060)	M7	1	--	--	--	--
M016001	1.267	(0.066)	0.617	(0.035)	0.117	(0.019)	--	--	M7	7	M7	7
M016801	1.190	(0.068)	1.650	(0.028)	0.044	(0.007)	--	--	M7	15	M7	15
M016901	1.629	(0.103)	1.100	(0.024)	0.160	(0.011)	--	--	M7	16	M7	16
M016902	1.404	(0.043)	1.430	(0.020)	0.000	(0.000)	--	--	M7	17	M7	17
M017101	1.320	(0.137)	1.366	(0.048)	0.212	(0.023)	--	--	--	--	M7	18
M017102	2.122	(0.246)	1.895	(0.032)	0.168	(0.009)	--	--	--	--	M7	19
M017301	1.507	(0.102)	2.246	(0.050)	0.000	(0.000)	--	--	--	--	M7	21
M018301	1.093	(0.035)	-0.064	(0.030)	0.071	(0.015)	M4	10	M4	10	M4	10
M018701	2.006	(0.096)	0.384	(0.018)	0.173	(0.009)	M4	14	M4	14	M4	14
M018801	1.684	(0.104)	0.822	(0.029)	0.279	(0.014)	--	--	M4	15	M4	15
M019301	2.254	(0.149)	1.056	(0.020)	0.192	(0.009)	--	--	M4	20	M4	20
M019401	3.815	(1.048)	2.110	(0.040)	0.285	(0.009)	--	--	--	--	M4	21
M019501	1.796	(0.179)	1.998	(0.035)	0.073	(0.008)	--	--	--	--	M4	22
M019701	0.859	(0.029)	-0.768	(0.020)	0.000	(0.000)	M5	1	M5	1	--	--
M020401	0.677	(0.017)	0.253	(0.018)	0.000	(0.000)	M5	8	M5	8	M5	8
M021201	1.228	(0.036)	0.634	(0.014)	0.000	(0.000)	--	--	M5	14	M5	14
M021601	1.519	(0.095)	2.340	(0.049)	0.000	(0.000)	--	--	--	--	M5	4
M021602	2.188	(0.126)	2.007	(0.028)	0.000	(0.000)	--	--	--	--	M5	5
M021702	0.998	(0.046)	0.906	(0.023)	0.000	(0.000)	--	--	--	--	M5	16
M022101	0.756	(0.048)	-2.204	(0.138)	0.255	(0.062)	M6	3	M6	3	--	--
M022401	1.836	(0.158)	-0.176	(0.041)	0.342	(0.020)	M6	6	M6	6	--	--
M023201	2.135	(0.176)	0.110	(0.026)	0.225	(0.013)	M6	15	M6	15	--	--
M024201	1.405	(0.095)	1.129	(0.036)	0.085	(0.020)	--	--	--	--	M6	6
M025001	2.220	(0.224)	0.507	(0.048)	0.219	(0.036)	--	--	--	--	M6	14
M025201	2.561	(0.577)	2.263	(0.060)	0.199	(0.008)	--	--	--	--	M6	16
M025401	1.357	(0.188)	1.366	(0.060)	0.304	(0.026)	--	--	--	--	M6	20
M025731	1.074	(0.108)	-1.333	(0.106)	0.254	(0.056)	M9	3	--	--	--	--
M025831	1.012	(0.063)	-1.221	(0.028)	0.000	(0.000)	M9	4	--	--	--	--
M026631	2.262	(0.230)	0.079	(0.036)	0.000	(0.000)	M9	12	--	--	--	--
M027131	1.118	(0.076)	-1.083	(0.089)	0.188	(0.053)	--	--	M9	2	--	--
M027531	1.235	(0.140)	0.182	(0.085)	0.376	(0.038)	--	--	M9	6	--	--
M027731	1.645	(0.159)	0.476	(0.038)	0.170	(0.023)	--	--	M9	8	--	--
M027931	1.373	(0.059)	0.357	(0.018)	0.000	(0.000)	--	--	M9	10	--	--
M028431	1.205	(0.058)	0.734	(0.026)	0.000	(0.000)	--	--	M9	15	--	--
M030231	1.190	(0.092)	1.255	(0.046)	0.113	(0.023)	--	--	--	--	M9	12
M030631	0.938	(0.119)	1.421	(0.078)	0.213	(0.034)	--	--	--	--	M9	16
M030831	1.984	(0.122)	1.848	(0.027)	0.000	(0.000)	--	--	--	--	M9	18
M030931	3.276	(0.942)	2.104	(0.049)	0.357	(0.011)	--	--	--	--	M9	19
M031031	2.100	(0.304)	1.953	(0.046)	0.188	(0.011)	--	--	--	--	M9	20
N255701	2.219	(0.277)	0.325	(0.064)	0.286	(0.049)	--	--	--	--	M3	9
N255701	1.804	(0.201)	0.795	(0.031)	0.125	(0.017)	--	--	M3	23	--	--
N256001	1.330	(0.056)	0.850	(0.018)	0.000	(0.000)	--	--	--	--	M3	18
N256101	1.131	(0.061)	-0.585	(0.031)	0.000	(0.000)	--	--	M3	7	--	--
N257101	1.295	(0.255)	2.261	(0.087)	0.319	(0.016)	--	--	--	--	M3	21

Table E-22 (continued)
1990 IRT Parameters, Mathematics Cross-sectional Samples
Algebra and Functions Subscale

							Age 9/Grade 4		Age 13/Grade 8		Age 17/Grade 12	
NAEP ID	A	S.E.	B	S.E.	C	S.E.	Block	Item	Block	Item	Block	Item
N262601	1.016	(0.102)	1.109	(0.074)	0.199	(0.035)	–	–	–	–	M3	17
N264701	2.084	(0.232)	0.590	(0.032)	0.173	(0.019)	–	–	M3	20	–	–
N270901	1.511	(0.109)	-2.229	(0.055)	0.000	(0.000)	M3	4	–	–	–	–
N271101	1.112	(0.075)	-1.176	(0.026)	0.000	(0.000)	M3	16	–	–	–	–

Table E-23
1990 IRT Parameters, Mathematics Cross-sectional Samples
Estimation Subscale

							Age 9/Grade 4		Age 13/Grade 8		Age 17/Grade 12	
NAEP ID	A	S.E.	B	S.E.	C	S.E.	Block	Item	Block	Item	Block	Item
M031101	1.257	(0.243)	-1.839	(0.147)	0.366	(0.068)	M10	11	--	--	--	--
M031201	0.333	(0.073)	-0.973	(0.414)	0.376	(0.067)	M10	12	--	--	--	--
M031301	1.155	(0.194)	-1.895	(0.144)	0.324	(0.065)	M10	13	--	--	--	--
M031401	0.983	(0.256)	-0.029	(0.136)	0.326	(0.044)	M10	14	--	--	--	--
M031402	1.389	(0.302)	-0.137	(0.077)	0.190	(0.032)	M10	15	--	--	--	--
M031501	1.052	(0.196)	-0.568	(0.125)	0.298	(0.051)	M10	16	--	--	--	--
M031601	0.866	(0.129)	-1.245	(0.152)	0.310	(0.061)	M10	17	--	--	--	--
M031701	0.695	(0.117)	-1.936	(0.216)	0.345	(0.067)	M10	18	--	--	--	--
M031801	1.483	(0.361)	-0.343	(0.101)	0.366	(0.039)	M10	19	--	--	--	--
M031901	1.432	(0.684)	0.535	(0.282)	0.578	(0.020)	M10	20	--	--	--	--
M032001	1.191	(0.260)	-0.207	(0.099)	0.254	(0.040)	M10	1	--	--	--	--
M032001	1.055	(0.103)	0.490	(0.091)	0.280	(0.044)	--	--	M10	1	M10	1
M032101	0.854	(0.336)	1.018	(0.481)	0.481	(0.026)	M10	2	M10	2	M10	2
M032201	1.027	(0.075)	-0.520	(0.100)	0.301	(0.040)	M10	3	M10	3	M10	3
M032301	0.948	(0.079)	0.075	(0.080)	0.267	(0.030)	M10	4	M10	4	M10	4
M032401	0.824	(0.083)	0.473	(0.083)	0.294	(0.028)	M10	5	M10	5	M10	5
M032501	0.634	(0.043)	0.080	(0.091)	0.147	(0.030)	M10	6	M10	6	M10	6
M032601	1.533	(0.134)	0.517	(0.045)	0.421	(0.016)	M10	7	M10	7	M10	7
M032701	0.789	(0.095)	-0.279	(0.184)	0.562	(0.041)	M10	8	M10	8	M10	8
M032801	2.748	(0.276)	0.989	(0.027)	0.311	(0.016)	M10	9	M10	9	M10	9
M032901	1.372	(0.090)	0.473	(0.035)	0.200	(0.015)	M10	10	M10	10	M10	10
M033001	1.062	(0.110)	1.106	(0.051)	0.137	(0.026)	--	--	M10	11	M10	11
M033101	0.902	(0.087)	0.417	(0.102)	0.245	(0.046)	--	--	M10	12	M10	12
M033201	1.016	(0.114)	0.230	(0.130)	0.390	(0.055)	--	--	M10	13	M10	13
M033301	1.938	(0.149)	0.776	(0.029)	0.142	(0.019)	--	--	M10	14	M10	14
M033401	1.133	(0.202)	1.262	(0.076)	0.418	(0.029)	--	--	M10	15	M10	15
M033501	1.057	(0.129)	0.637	(0.100)	0.349	(0.044)	--	--	M10	16	M10	16
M033601	1.234	(0.176)	1.532	(0.058)	0.176	(0.020)	--	--	M10	17	M10	17
M033701	1.605	(0.182)	1.119	(0.039)	0.247	(0.021)	--	--	M10	18	M10	18
M033801	1.671	(0.142)	0.674	(0.044)	0.249	(0.027)	--	--	M10	19	M10	19
M033901	1.031	(0.141)	1.078	(0.071)	0.257	(0.034)	--	--	M10	20	M10	20
M034001	1.328	(0.252)	1.675	(0.080)	0.261	(0.019)	--	--	M10	21	M10	21
M034101	1.659	(0.167)	1.236	(0.031)	0.120	(0.015)	--	--	M10	22	M10	22
M034601	1.345	(0.134)	0.216	(0.097)	0.413	(0.047)	--	--	M11	1	M11	1
M034602	1.353	(0.114)	0.235	(0.083)	0.352	(0.044)	--	--	M11	2	M11	2
M034603	2.460	(0.338)	0.745	(0.046)	0.549	(0.021)	--	--	M11	3	M11	3
M034604	2.796	(0.375)	0.808	(0.039)	0.541	(0.019)	--	--	M11	4	M11	4
M034701	1.574	(0.118)	-0.157	(0.071)	0.261	(0.050)	--	--	M11	5	M11	5
M034702	1.735	(0.154)	0.568	(0.046)	0.240	(0.030)	--	--	M11	6	M11	6
M034703	1.575	(0.177)	0.637	(0.059)	0.346	(0.032)	--	--	M11	7	M11	7
M034704	1.445	(0.133)	0.073	(0.086)	0.336	(0.050)	--	--	M11	8	M11	8
M034801	1.577	(0.178)	-0.029	(0.103)	0.473	(0.053)	--	--	M11	9	M11	9
M034901	1.852	(0.212)	0.851	(0.046)	0.408	(0.023)	--	--	M11	10	M11	10
M035001	2.160	(0.273)	0.834	(0.041)	0.428	(0.022)	--	--	M11	11	M11	11
M035101	2.143	(0.199)	0.618	(0.041)	0.339	(0.025)	--	--	M11	12	M11	12
M035201	1.435	(0.126)	0.340	(0.068)	0.273	(0.040)	--	--	M11	13	M11	13
M035301	1.386	(0.125)	-0.039	(0.093)	0.344	(0.054)	--	--	M11	14	M11	14
M035401	1.735	(0.220)	1.130	(0.040)	0.313	(0.020)	--	--	M11	15	M11	15
M035501	0.834	(0.114)	0.768	(0.123)	0.318	(0.048)	--	--	M11	16	M11	16
M035601	3.490	(0.330)	1.161	(0.016)	0.092	(0.009)	--	--	M11	17	M11	17
M035701	1.428	(0.136)	0.515	(0.063)	0.272	(0.037)	--	--	M11	18	M11	18
M035801	1.596	(0.191)	0.913	(0.051)	0.369	(0.026)	--	--	M11	19	M11	19
M035901	1.280	(0.123)	0.376	(0.086)	0.350	(0.043)	--	--	M11	20	M11	20

Table E-23 (continued)
1990 IRT Parameters, Mathematics Cross-sectional Samples
Estimation Subscale

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 9/Grade 4 Block	Age 9/Grade 4 Item	Age 13/Grade 8 Block	Age 13/Grade 8 Item	Age 17/Grade 12 Block	Age 17/Grade 12 Item
M036001	0.854	(0.091)	0.637	(0.103)	0.244	(0.044)	--	--	M11	21	M11	21
M036101	2.640	(0.321)	1.110	(0.027)	0.307	(0.015)	--	--	M11	22	M11	22
M036201	1.291	(0.114)	0.281	(0.082)	0.295	(0.045)	--	--	M11	23	M11	23
M036301	1.698	(0.195)	0.960	(0.046)	0.365	(0.023)	--	--	M11	24	M11	24

Table E-24
1990 IRT Parameters, Science Trend Sample, Age 9

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 9 Block	Age 9 Item
N400001	0.638	(0.068)	-1.118	(0.224)	0.317	(0.064)	S1	6
N400101	0.709	(0.255)	2.213	(0.415)	0.540	(0.023)	S1	15
N400102	0.743	(0.209)	1.654	(0.223)	0.464	(0.027)	S1	16
N400301	1.268	(0.237)	0.043	(0.096)	0.399	(0.035)	S1	8
N400401	1.325	(0.197)	-1.024	(0.187)	0.492	(0.058)	S1	9
N400402	1.833	(0.241)	-0.628	(0.095)	0.292	(0.040)	S1	10
N400403	0.579	(0.068)	-1.720	(0.327)	0.498	(0.068)	S1	11
N400404	1.286	(0.163)	-0.489	(0.117)	0.334	(0.042)	S1	12
N400405	1.113	(0.177)	-0.563	(0.172)	0.457	(0.050)	S1	13
N400501	0.743	(0.147)	0.625	(0.119)	0.372	(0.038)	S1	14
N400601	0.796	(0.120)	-0.054	(0.138)	0.280	(0.046)	S1	17
N400701	0.967	(0.140)	0.221	(0.082)	0.244	(0.034)	S1	18
N400901	0.328	(0.086)	2.443	(0.403)	0.309	(0.043)	S1	19
N401001	0.685	(0.115)	0.711	(0.100)	0.256	(0.035)	S1	20
N401101	1.085	(0.310)	1.591	(0.178)	0.430	(0.020)	S1	21
N401201	0.796	(0.189)	2.019	(0.219)	0.239	(0.020)	S1	22
N401301	0.713	(0.146)	1.167	(0.115)	0.303	(0.031)	S1	23
N401501	0.285	(0.061)	0.819	(0.458)	0.410	(0.059)	S2	1
N401601	0.586	(0.060)	-1.350	(0.216)	0.257	(0.062)	S2	2
N401702	0.448	(0.135)	1.683	(0.377)	0.583	(0.037)	S2	4
N401703	0.363	(0.106)	1.986	(0.461)	0.530	(0.042)	S2	5
N401801	1.000	(0.256)	0.285	(0.141)	0.534	(0.039)	S2	6
N401802	0.676	(0.136)	-0.361	(0.284)	0.552	(0.055)	S2	7
N401803	0.613	(0.142)	0.353	(0.250)	0.556	(0.047)	S2	8
N401804	0.595	(0.167)	1.527	(0.241)	0.483	(0.036)	S2	9
N401901	0.639	(0.153)	1.500	(0.183)	0.352	(0.034)	S2	10
N402001	1.107	(0.178)	-0.721	(0.191)	0.435	(0.058)	S2	11
N402002	1.337	(0.238)	-0.766	(0.182)	0.447	(0.058)	S2	12
N402005	0.997	(0.237)	-0.056	(0.176)	0.478	(0.049)	S2	15
N402101	0.642	(0.096)	-0.038	(0.171)	0.260	(0.053)	S2	16
N402201	0.310	(0.056)	0.690	(0.348)	0.303	(0.058)	S2	17
N402401	0.327	(0.098)	3.491	(0.725)	0.314	(0.036)	S2	18
N402501	1.559	(0.467)	1.643	(0.128)	0.259	(0.015)	S2	19
N402602	0.401	(0.000)	-0.452	(0.285)	0.505	(0.049)	S2	21
N402701	0.456	(0.102)	2.127	(0.262)	0.231	(0.036)	S2	23
N402801	1.375	(0.345)	1.860	(0.139)	0.177	(0.013)	S2	24
N402901	0.562	(0.207)	4.257	(1.037)	0.203	(0.014)	S2	25
N403001	0.339	(0.054)	-6.408	(0.982)	0.296	(0.065)	S3	12
N403101	0.626	(0.062)	-3.565	(0.288)	0.283	(0.064)	S3	13
N403201	0.489	(0.042)	-2.507	(0.253)	0.244	(0.061)	S3	14
N403202	0.342	(0.040)	-1.009	(0.325)	0.274	(0.060)	S3	15
N403301	0.635	(0.063)	-0.975	(0.189)	0.261	(0.057)	S3	16
N403401	0.426	(0.086)	0.518	(0.276)	0.392	(0.055)	S3	17
N403501	0.752	(0.142)	0.312	(0.146)	0.421	(0.042)	S3	18
N403502	0.588	(0.063)	-1.816	(0.302)	0.459	(0.068)	S3	19
N403503	0.477	(0.099)	0.281	(0.291)	0.459	(0.055)	S3	20
N403601	0.816	(0.127)	0.604	(0.080)	0.241	(0.033)	S3	21
N403701	3.290	(0.000)	-0.087	(0.026)	0.318	(0.019)	S3	22
N403702	3.150	(0.000)	-0.265	(0.037)	0.422	(0.023)	S3	23
N403703	2.988	(0.438)	-0.100	(0.052)	0.331	(0.026)	S3	24
N403801	0.381	(0.103)	1.686	(0.335)	0.458	(0.045)	S3	25
N403803	0.518	(0.073)	-0.787	(0.313)	0.453	(0.062)	S3	27
N403804	0.516	(0.082)	-0.357	(0.287)	0.426	(0.059)	S3	28
N403901	0.767	(0.088)	-0.171	(0.126)	0.229	(0.045)	S3	29
N404001	0.233	(0.047)	1.730	(0.432)	0.254	(0.053)	S3	30
N404201	0.530	(0.112)	1.351	(0.146)	0.249	(0.038)	S3	31

Table E-25
1990 IRT Parameters, Science Trend Sample, Age 13

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 13 Block	Age 13 Item
N400201	0.410	(0.038)	-1.648	(0.257)	0.240	(0.060)	S1	16
N401201	0.616	(0.063)	0.222	(0.143)	0.255	(0.044)	S1	28
N404501	0.981	(0.088)	-2.296	(0.140)	0.198	(0.056)	S1	12
N404601	0.390	(0.049)	-0.223	(0.282)	0.274	(0.061)	S1	13
N404701	0.552	(0.043)	-1.687	(0.187)	0.226	(0.057)	S1	14
N404702	0.538	(0.056)	-0.020	(0.174)	0.235	(0.051)	S1	15
N404801	1.214	(0.194)	-1.575	(0.184)	0.468	(0.067)	S1	20
N404802	2.083	(0.326)	-0.323	(0.069)	0.385	(0.034)	S1	21
N404803	1.087	(0.131)	0.329	(0.086)	0.349	(0.032)	S1	22
N404901	0.676	(0.051)	-0.684	(0.138)	0.225	(0.050)	S1	17
N405001	0.518	(0.065)	0.412	(0.186)	0.281	(0.049)	S1	23
N405101	0.995	(0.123)	0.916	(0.059)	0.216	(0.023)	S1	24
N405201	0.267	(0.036)	-0.061	(0.367)	0.233	(0.060)	S1	25
N405301	0.794	(0.122)	1.212	(0.080)	0.222	(0.027)	S1	26
N405401	1.208	(0.159)	1.006	(0.047)	0.139	(0.018)	S1	27
N405501	0.833	(0.096)	0.142	(0.114)	0.279	(0.042)	S1	29
N405601	0.241	(0.043)	1.573	(0.409)	0.244	(0.056)	S1	30
N405701	1.210	(0.132)	0.622	(0.050)	0.186	(0.023)	S1	31
N405801	0.939	(0.142)	1.250	(0.068)	0.234	(0.021)	S1	32
N405901	0.636	(0.123)	2.002	(0.152)	0.185	(0.025)	S1	33
N406001	1.094	(0.270)	2.769	(0.238)	0.157	(0.009)	S1	34
N406101	0.849	(0.227)	3.118	(0.346)	0.202	(0.011)	S1	35
N406201	0.549	(0.117)	3.151	(0.343)	0.101	(0.017)	S1	36
N406301	0.283	(0.048)	-1.346	(0.573)	0.487	(0.067)	S2	10
N406302	0.369	(0.055)	-0.506	(0.412)	0.454	(0.064)	S2	11
N406303	0.955	(0.256)	1.520	(0.129)	0.468	(0.022)	S2	12
N406304	0.579	(0.150)	1.487	(0.205)	0.478	(0.036)	S2	13
N406401	0.613	(0.109)	0.032	(0.271)	0.522	(0.056)	S2	14
N406402	0.945	(0.148)	0.222	(0.135)	0.422	(0.043)	S2	15
N406403	0.808	(0.107)	-1.278	(0.233)	0.488	(0.065)	S2	16
N406404	1.028	(0.164)	-0.197	(0.167)	0.527	(0.048)	S2	17
N406405	0.742	(0.094)	-0.393	(0.204)	0.445	(0.055)	S2	18
N406501	0.549	(0.073)	0.703	(0.137)	0.189	(0.043)	S2	19
N406601	0.484	(0.042)	-0.885	(0.194)	0.206	(0.055)	S2	20
N406701	0.629	(0.082)	0.241	(0.152)	0.275	(0.047)	S2	21
N406801	0.936	(0.085)	-1.585	(0.178)	0.429	(0.065)	S2	22
N406802	0.451	(0.092)	0.925	(0.286)	0.502	(0.047)	S2	23
N406803	0.794	(0.089)	-0.670	(0.178)	0.384	(0.057)	S2	24
N406804	0.826	(0.076)	-1.041	(0.175)	0.398	(0.059)	S2	25
N406805	0.974	(0.315)	1.761	(0.189)	0.587	(0.017)	S2	26
N406806	0.342	(0.061)	0.502	(0.399)	0.436	(0.059)	S2	27
N406901	0.528	(0.061)	-0.048	(0.191)	0.254	(0.053)	S2	28
N407001	0.292	(0.041)	0.311	(0.312)	0.212	(0.056)	S2	29
N407101	0.809	(0.160)	2.112	(0.158)	0.141	(0.018)	S2	30
N407201	0.458	(0.063)	0.750	(0.193)	0.264	(0.047)	S2	31
N407301	0.389	(0.071)	1.445	(0.212)	0.258	(0.046)	S2	32
N407302	0.578	(0.153)	2.019	(0.224)	0.381	(0.031)	S2	33
N407601	0.575	(0.100)	1.533	(0.122)	0.205	(0.033)	S2	35
N407701	0.533	(0.079)	1.303	(0.122)	0.169	(0.035)	S2	37
N407801	0.870	(0.188)	1.882	(0.145)	0.234	(0.018)	S2	38
N407901	0.362	(0.050)	0.823	(0.232)	0.201	(0.050)	S2	39
N408001	1.164	(0.157)	0.989	(0.049)	0.222	(0.018)	S2	34
N408201	0.541	(0.125)	3.074	(0.346)	0.173	(0.020)	S2	40
N408301	1.198	(0.175)	0.926	(0.061)	0.342	(0.021)	S3	10

Table E-25 (continued)
1990 IRT Parameters, Science Trend Sample, Age 13

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 13 Block	Age 13 Item
N408302	0.809	(0.080)	-1.148	(0.193)	0.422	(0.062)	S3	11
N408303	0.730	(0.076)	-1.383	(0.232)	0.457	(0.066)	S3	12
N408304	1.000	(0.114)	-1.039	(0.173)	0.452	(0.060)	S3	13
N408401	0.291	(0.033)	-1.179	(0.357)	0.252	(0.061)	S3	14
N408501	0.679	(0.057)	-0.770	(0.153)	0.234	(0.053)	S3	15
N408502	0.461	(0.074)	1.418	(0.149)	0.191	(0.039)	S3	16
N408601	0.526	(0.039)	-0.864	(0.150)	0.149	(0.048)	S3	17
N408701	0.335	(0.049)	0.363	(0.309)	0.271	(0.058)	S3	18
N408801	0.201	(0.034)	0.896	(0.508)	0.268	(0.059)	S3	19
N408901	0.863	(0.139)	0.432	(0.130)	0.487	(0.035)	S3	20
N408902	0.855	(0.089)	-1.856	(0.212)	0.469	(0.069)	S3	21
N408903	1.014	(0.151)	0.521	(0.093)	0.423	(0.030)	S3	22
N408904	0.551	(0.096)	0.929	(0.187)	0.422	(0.042)	S3	23
N409001	0.597	(0.057)	-0.195	(0.147)	0.200	(0.048)	S3	24
N409101	0.713	(0.053)	-1.354	(0.160)	0.271	(0.059)	S3	25
N409102	0.679	(0.095)	0.479	(0.128)	0.304	(0.041)	S3	26
N409103	1.489	(0.345)	1.528	(0.083)	0.362	(0.013)	S3	27
N409201	0.360	(0.055)	0.627	(0.284)	0.290	(0.056)	S3	28
N409301	0.828	(0.082)	-0.075	(0.101)	0.186	(0.041)	S3	29
N409501	0.876	(0.134)	1.734	(0.099)	0.151	(0.016)	S3	33
N409601	1.102	(0.195)	1.367	(0.076)	0.301	(0.018)	S3	34
N409701	0.746	(0.134)	2.113	(0.151)	0.155	(0.018)	S3	35

Table E-26
1990 IRT Parameters, Science Trend Sample, Age 17

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 17 Block	Age 17 Item
N400201	0.493	(0.073)	-2.508	(0.370)	0.221	(0.058)	S1	12
N401201	0.998	(0.139)	-0.053	(0.127)	0.288	(0.048)	S1	30
N404601	0.406	(0.065)	-1.520	(0.336)	0.238	(0.061)	S1	13
N405001	0.460	(0.066)	-0.558	(0.227)	0.207	(0.054)	S1	29
N405101	1.043	(0.163)	0.018	(0.131)	0.287	(0.050)	S3	14
N405201	0.464	(0.068)	-0.409	(0.222)	0.190	(0.055)	S1	31
N405401	0.843	(0.137)	0.481	(0.123)	0.195	(0.044)	S3	19
N405501	0.604	(0.087)	-0.739	(0.209)	0.234	(0.059)	S3	21
N406001	0.713	(0.218)	2.053	(0.295)	0.251	(0.031)	S1	33
N406101	0.[illegible]70	(0.213)	2.091	(0.272)	0.159	(0.026)	S1	35
N406201	1.277	(0.331)	1.768	(0.148)	0.125	(0.017)	S1	37
N406301	0.297	(0.067)	-1.179	(0.552)	0.441	(0.068)	S1	21
N406302	0.427	(0.076)	-0.869	(0.365)	0.396	(0.066)	S1	22
N406303	0.556	(0.102)	-0.223	(0.284)	0.412	(0.064)	S1	23
N406304	0.439	(0.082)	-0.724	(0.354)	0.391	(0.066)	S1	24
N406401	0.785	(0.126)	-0.527	(0.204)	0.374	(0.061)	S2	10
N406402	0.891	(0.142)	-0.641	(0.196)	0.400	(0.062)	S2	11
N406403	0.883	(0.155)	-1.737	(0.265)	0.416	(0.069)	S2	12
N406404	1.067	(0.145)	-0.973	(0.174)	0.400	(0.064)	S2	13
N406405	0.971	(0.148)	-1.214	(0.212)	0.410	(0.067)	S2	14
N406601	0.425	(0.068)	-0.935	(0.262)	0.191	(0.057)	S1	28
N406801	0.839	(0.128)	-2.132	(0.282)	0.397	(0.068)	S2	16
N406802	0.359	(0.086)	0.827	(0.422)	0.418	(0.060)	S2	17
N406803	0.660	(0.085)	-1.231	(0.241)	0.372	(0.064)	S2	18
N406804	0.649	(0.098)	-2.014	(0.319)	0.408	(0.068)	S2	19
N406805	0.611	(0.115)	0.102	(0.251)	0.421	(0.059)	S2	20
N406806	0.296	(0.069)	0.470	(0.505)	0.422	(0.063)	S2	21
N406901	0.538	(0.084)	-0.516	(0.231)	0.256	(0.061)	S2	27
N407001	0.313	(0.055)	-0.789	(0.331)	0.189	(0.056)	S2	33
N407101	0.811	(0.144)	1.304	(0.126)	0.140	(0.032)	S2	38
N407201	0.511	(0.071)	0.047	(0.184)	0.174	(0.050)	S2	32
N407301	0.312	(0.063)	1.055	(0.349)	0.221	(0.054)	S2	36
N407302	0.413	(0.098)	1.129	(0.317)	0.299	(0.057)	S2	37
N407401	0.427	(0.083)	-0.816	(0.381)	0.412	(0.067)	S2	28
N407403	0.709	(0.125)	-0.266	(0.231)	0.406	(0.062)	S2	30
N407404	0.777	(0.128)	-1.638	(0.273)	0.423	(0.068)	S2	31
N407701	0.476	(0.082)	0.862	(0.194)	0.155	(0.045)	S2	35
N408101	0.804	(0.206)	1.832	(0.203)	0.171	(0.029)	S1	38
N408301	0.795	(0.111)	-0.610	(0.199)	0.365	(0.062)	S3	10
N408302	0.788	(0.142)	-2.013	(0.319)	0.398	(0.068)	S3	11
N408303	0.557	(0.114)	-3.004	(0.566)	0.409	(0.069)	S3	12
N408304	0.754	(0.136)	-2.296	(0.358)	0.406	(0.069)	S3	13
N408601	0.470	(0.060)	-1.580	(0.243)	0.146	(0.049)	S1	19
N408801	0.794	(0.128)	-0.245	(0.173)	0.273	(0.059)	S3	24
N408901	0.838	(0.127)	-1.247	(0.231)	0.432	(0.068)	S3	15
N408902	1.127	(0.256)	-2.288	(0.322)	0.397	(0.068)	S3	16
N408903	0.634	(0.103)	-0.470	(0.244)	0.378	(0.063)	S3	17
N408904	0.402	(0.074)	-0.656	(0.385)	0.407	(0.067)	S3	18
N409301	0.702	(0.090)	-1.635	(0.205)	0.177	(0.054)	S1	20
N409501	0.862	(0.153)	1.174	(0.108)	0.121	(0.029)	S1	34
N409901	0.759	(0.106)	-1.066	(0.189)	0.243	(0.059)	S1	18
N410003	0.305	(0.065)	-4.157	(0.949)	0.419	(0.069)	S1	16
N410004	0.416	(0.077)	-1.713	(0.438)	0.423	(0.069)	S1	17
N410101	0.712	(0.138)	-0.777	(0.275)	0.487	(0.065)	S1	25

Table E-26 (continued)
1990 IRT Parameters, Science Trend Sample, Age 17

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 17 Block	Age 17 Item
N410102	0.280	(0.061)	-1.386	(0.580)	0.428	(0.068)	S1	26
N410103	0.374	(0.067)	-2.099	(0.498)	0.407	(0.068)	S1	27
N410201	0.828	(0.211)	1.476	(0.166)	0.207	(0.035)	S1	32
N410401	0.297	(0.059)	-0.323	(0.393)	0.275	(0.063)	S2	15
N410501	0.434	(0.057)	-0.860	(0.217)	0.150	(0.047)	S2	22
N410601	1.613	(0.403)	1.412	(0.089)	0.170	(0.018)	S2	23
N410602	0.563	(0.098)	-2.340	(0.407)	0.414	(0.069)	S2	24
N410603	1.411	(0.368)	1.060	(0.101)	0.383	(0.029)	S2	25
N410604	0.407	(0.079)	-2.980	(0.611)	0.411	(0.069)	S2	26
N410701	0.834	(0.149)	0.860	(0.122)	0.216	(0.039)	S2	34
N410801	0.515	(0.135)	1.754	(0.290)	0.230	(0.044)	S2	39
N410901	1.208	(0.204)	1.138	(0.078)	0.115	(0.021)	S2	40
N411001	1.096	(0.261)	1.562	(0.131)	0.142	(0.022)	S2	41
N411101	0.508	(0.076)	-0.183	(0.193)	0.172	(0.052)	S3	22
N411201	0.717	(0.112)	0.217	(0.154)	0.224	(0.050)	S3	23
N411301	0.769	(0.284)	2.844	(0.538)	0.141	(0.023)	S3	20
N411401	2.192	(0.491)	0.449	(0.062)	0.211	(0.030)	S3	25
N411501	0.861	(0.179)	1.261	(0.123)	0.155	(0.032)	S3	26
N411502	0.705	(0.109)	-1.311	(0.223)	0.253	(0.061)	S3	27
N411601	1.279	(0.256)	0.914	(0.083)	0.193	(0.031)	S3	28
N411701	1.083	(0.250)	1.113	(0.102)	0.197	(0.032)	S3	29
N411801	1.906	(0.320)	0.371	(0.063)	0.164	(0.031)	S3	30
N411901	0.900	(0.197)	1.097	(0.117)	0.173	(0.037)	S3	31
N412001	1.198	(0.456)	1.566	(0.175)	0.280	(0.031)	S3	32

Table E-27
1990 IRT Parameters, Science Cross-sectional Samples
Life Sciences Subscale

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 9/Grade 4 Block	Age 9/Grade 4 Item	Age 13/Grade 8 Block	Age 13/Grade 8 Item	Age 17/Grade 12 Block	Age 17/Grade 12 Item
K010601	1.034	(0.043)	-0.173	(0.046)	0.221	(0.020)	S4	6	S4	6	S4	6
K010901	0.941	(0.042)	0.212	(0.043)	0.211	(0.017)	S4	9	S4	9	S4	9
K011501	1.115	(0.058)	0.555	(0.032)	0.268	(0.013)	S4	15	S4	15	S4	15
K011701	1.449	(0.061)	0.280	(0.025)	0.244	(0.011)	S4	17	S4	17	S4	17
K012101	2.289	(0.177)	1.433	(0.020)	0.152	(0.004)	S4	20	S4	20	S4	20
K012102	1.705	(0.087)	0.923	(0.016)	0.178	(0.007)	S4	21	S4	21	S4	21
K012201	1.178	(0.083)	1.075	(0.031)	0.197	(0.016)	--	--	S4	22	S4	22
K012301	1.385	(0.100)	1.029	(0.029)	0.261	(0.014)	--	--	S4	23	S4	23
K012401	1.412	(0.074)	0.341	(0.041)	0.230	(0.024)	--	--	S4	24	S4	24
K013701	0.898	(0.107)	-0.830	(0.108)	0.269	(0.045)	S5	7	--	--	--	--
K013701	0.838	(0.068)	-0.086	(0.105)	0.212	(0.048)	--	--	S5	7	--	--
K013801	1.474	(0.082)	0.039	(0.030)	0.228	(0.014)	S5	8	S5	8	--	--
K013802	2.366	(0.142)	-0.085	(0.021)	0.225	(0.011)	S5	9	S5	9	--	--
K014101	1.747	(0.093)	0.011	(0.024)	0.194	(0.012)	S5	11	S5	11	--	--
K014301	1.198	(0.071)	0.036	(0.039)	0.250	(0.017)	S5	13	S5	13	--	--
K015101	1.882	(0.158)	0.014	(0.044)	0.210	(0.032)	--	--	S5	20	--	--
K015201	1.247	(0.139)	0.749	(0.046)	0.236	(0.025)	--	--	S5	21	--	--
K015701	0.691	(0.066)	0.341	(0.148)	0.243	(0.057)	--	--	--	--	S5	1
K016901	1.832	(0.244)	1.478	(0.032)	0.219	(0.015)	--	--	--	--	S5	15
K017201	1.546	(0.153)	1.169	(0.032)	0.159	(0.020)	--	--	--	--	S5	17
K017202	1.652	(0.278)	1.952	(0.067)	0.178	(0.012)	--	--	--	--	S5	18
K017501	1.771	(0.329)	2.147	(0.091)	0.138	(0.009)	--	--	--	--	S5	21
K017801	0.902	(0.043)	-1.843	(0.091)	0.180	(0.047)	S6	1	S6	1	--	--
K017901	0.474	(0.028)	-1.613	(0.175)	0.203	(0.052)	S6	2	S6	2	--	--
K018601	1.689	(0.106)	0.293	(0.024)	0.206	(0.011)	S6	8	S6	8	--	--
K018602	1.245	(0.082)	0.457	(0.031)	0.209	(0.013)	S6	9	S6	9	--	--
K019701	1.549	(0.129)	0.770	(0.043)	0.156	(0.028)	--	--	--	--	S6	3
K020101	1.173	(0.194)	1.527	(0.055)	0.317	(0.025)	--	--	--	--	S6	6
K020401	1.294	(0.184)	1.668	(0.048)	0.207	(0.018)	--	--	--	--	S6	9
K020501	1.116	(0.280)	2.456	(0.185)	0.289	(0.015)	--	--	--	--	S6	10
K021301	0.653	(0.042)	0.460	(0.039)	0.000	(0.000)	--	--	--	--	S6	17
K021401	1.114	(0.099)	-1.555	(0.093)	0.197	(0.047)	S7	1	--	--	--	--
K021501	0.831	(0.059)	-1.906	(0.112)	0.181	(0.048)	S7	2	--	--	--	--
K022101	0.789	(0.075)	-0.954	(0.096)	0.183	(0.041)	S7	7	--	--	--	--
K022201	1.631	(0.190)	-1.564	(0.084)	0.219	(0.047)	S7	8	--	--	--	--
K022901	0.728	(0.086)	-0.179	(0.090)	0.213	(0.034)	S7	15	--	--	--	--
K023701	0.720	(0.044)	0.325	(0.086)	0.147	(0.037)	--	--	S7	2	S7	2
K023901	1.232	(0.068)	-0.068	(0.064)	0.223	(0.039)	--	--	S7	4	S7	4
K023902	2.010	(0.135)	0.685	(0.026)	0.305	(0.014)	--	--	S7	5	S7	5
K023903	2.349	(0.162)	1.094	(0.014)	0.143	(0.007)	--	--	S7	6	S7	6
K024401	1.079	(0.074)	0.750	(0.046)	0.205	(0.023)	--	--	S7	10	S7	10
K024901	1.764	(0.127)	1.038	(0.023)	0.275	(0.011)	--	--	S7	16	S7	16
K025701	0.943	(0.120)	0.855	(0.101)	0.243	(0.045)	--	--	--	--	S7	25
K026601	0.575	(0.017)	1.280	(0.034)	0.000	(0.000)	S8	7	S8	7	S8	7
K026901	1.372	(0.030)	0.482	(0.011)	0.000	(0.000)	S8	10	S8	10	S8	10
K027201	0.721	(0.026)	1.046	(0.025)	0.000	(0.000)	--	--	S8	12	S8	12
K027401	0.909	(0.036)	1.675	(0.037)	0.000	(0.000)	--	--	S8	14	S8	14
K027801	1.021	(0.051)	-0.794	(0.064)	0.197	(0.032)	S9	2	S9	2	--	--
K027901	0.884	(0.044)	-0.653	(0.069)	0.167	(0.032)	S9	3	S9	3	--	--
K028501	1.014	(0.073)	0.410	(0.043)	0.238	(0.017)	S9	8	S9	8	--	--
K029101	1.034	(0.067)	-0.444	(0.087)	0.207	(0.049)	--	--	S9	14	--	--
K029201	0.537	(0.096)	1.374	(0.141)	0.242	(0.043)	--	--	S9	15	--	--
K029501	1.204	(0.082)	0.041	(0.080)	0.206	(0.049)	--	--	--	--	S9	1

Table E-27 (continued)
1990 IRT Parameters, Science Cross-sectional Samples
Life Sciences Subscale

							Age 9/Grade 4		Age 13/Grade 8		Age 17/Grade 12	
NAEP ID	A	S.E.	B	S.E.	C	S.E.	Block	Item	Block	Item	Block	Item
K029801	1.289	(0.090)	0.057	(0.075)	0.192	(0.048)	--	--	--	--	S9	4
K029802	0.669	(0.064)	0.837	(0.106)	0.158	(0.042)	--	--	--	--	S9	5
K030701	0.925	(0.050)	0.693	(0.026)	0.000	(0.000)	--	--	--	--	S9	13
N400201	0.859	(0.038)	-0.366	(0.064)	0.159	(0.029)	S3	1	S3	1	S3	1
N400601	1.130	(0.146)	-0.714	(0.083)	0.262	(0.042)	S3	8	--	--	--	--
N400701	1.308	(0.160)	-0.456	(0.056)	0.208	(0.031)	S3	12	--	--	--	--
N401201	0.981	(0.055)	0.581	(0.037)	0.247	(0.015)	S3	2	S3	2	S3	2
N401301	0.913	(0.140)	-0.003	(0.078)	0.232	(0.032)	S3	11	--	--	--	--
N404601	0.747	(0.053)	0.267	(0.103)	0.203	(0.043)	--	--	S3	5	S3	5
N404701	1.610	(0.125)	-0.023	(0.051)	0.208	(0.035)	--	--	S3	21	--	--
N404702	0.985	(0.084)	0.280	(0.073)	0.162	(0.038)	--	--	S3	22	--	--
N405001	0.602	(0.042)	0.332	(0.113)	0.161	(0.042)	--	--	S3	4	S3	4
N405701	1.336	(0.132)	0.814	(0.037)	0.141	(0.021)	--	--	S3	18	--	--
N405801	1.776	(0.215)	0.891	(0.032)	0.250	(0.016)	--	--	S3	19	--	--
N408801	0.845	(0.067)	0.604	(0.078)	0.216	(0.034)	--	--	S3	8	S3	8
N409901	1.574	(0.105)	0.209	(0.056)	0.167	(0.041)	--	--	--	--	S3	14
N411201	0.969	(0.114)	1.205	(0.064)	0.183	(0.033)	--	--	--	--	S3	22
N411502	1.303	(0.095)	0.051	(0.078)	0.212	(0.051)	--	--	--	--	S3	15

Table E-28
1990 IRT Parameters, Science Cross-sectional Samples
Physical Sciences Subscale

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 9/Grade 4 Block	Age 9/Grade 4 Item	Age 13/Grade 8 Block	Age 13/Grade 8 Item	Age 17/Grade 12 Block	Age 17/Grade 12 Item
K010201	1.553	(0.158)	1.184	(0.032)	0.383	(0.012)	--	--	S4	2	S4	2
K010401	0.843	(0.031)	-0.769	(0.065)	0.156	(0.030)	S4	4	S4	4	S4	4
K010501	1.673	(0.073)	-0.381	(0.029)	0.195	(0.016)	S4	5	S4	5	S4	5
K011401	1.280	(0.066)	-0.013	(0.035)	0.300	(0.016)	S4	14	S4	14	S4	14
K012501	1.927	(0.187)	1.547	(0.029)	0.200	(0.007)	--	--	S4	25	S4	25
K012701	0.796	(0.138)	2.292	(0.108)	0.277	(0.022)	--	--	--	--	S4	27
K012801	0.897	(0.156)	2.563	(0.119)	0.246	(0.016)	--	--	--	--	S4	28
K013101	0.734	(0.039)	-0.867	(0.094)	0.180	(0.037)	S5	1	S5	1	--	--
K013401	1.191	(0.068)	-0.849	(0.063)	0.239	(0.031)	S5	4	S5	4	--	--
K014201	1.454	(0.104)	0.399	(0.027)	0.293	(0.011)	S5	12	S5	12	--	--
K014401	0.957	(0.075)	0.128	(0.052)	0.286	(0.021)	S5	14	S5	14	--	--
K014801	0.729	(0.071)	1.058	(0.050)	0.146	(0.016)	S5	18	S5	18	--	--
K015301	1.804	(0.168)	0.506	(0.029)	0.174	(0.018)	--	--	S5	22	--	--
K015801	1.242	(0.081)	-0.010	(0.071)	0.240	(0.043)	--	--	--	--	S5	2
K015901	0.668	(0.053)	0.667	(0.108)	0.186	(0.041)	--	--	--	--	S5	3
K016101	0.716	(0.048)	0.285	(0.095)	0.146	(0.040)	--	--	--	--	S5	5
K016201	1.373	(0.109)	0.903	(0.038)	0.138	(0.022)	--	--	--	--	S5	6
K016501	0.905	(0.068)	0.133	(0.099)	0.230	(0.047)	--	--	--	--	S5	9
K016502	1.076	(0.087)	1.002	(0.056)	0.205	(0.026)	--	--	--	--	S5	10
K016601	0.831	(0.107)	2.480	(0.100)	0.086	(0.014)	--	--	--	--	S5	11
K017101	0.906	(0.101)	2.112	(0.068)	0.170	(0.016)	--	--	--	--	S5	16
K017701	1.151	(0.112)	1.458	(0.046)	0.180	(0.019)	--	--	--	--	S5	23
K018401	1.652	(0.092)	-0.068	(0.026)	0.201	(0.013)	S6	6	S6	6	--	--
K018501	0.595	(0.062)	0.153	(0.116)	0.281	(0.036)	S6	7	S6	7	--	--
K018802	1.040	(0.031)	0.144	(0.015)	0.000	(0.000)	S6	12	S6	12	--	--
K018803	1.154	(0.034)	0.140	(0.014)	0.000	(0.000)	S6	13	S6	13	--	--
K018901	0.901	(0.080)	-0.428	(0.121)	0.267	(0.054)	--	--	S6	14	--	--
K019101	0.755	(0.090)	0.780	(0.075)	0.171	(0.033)	--	--	S6	15	--	--
K019301	1.482	(0.295)	1.458	(0.090)	0.274	(0.013)	--	--	S6	17	--	--
K019501	0.724	(0.060)	0.674	(0.109)	0.185	(0.044)	--	--	--	--	S6	1
K019601	0.573	(0.051)	0.156	(0.164)	0.212	(0.056)	--	--	--	--	S6	2
K019901	0.523	(0.046)	0.642	(0.146)	0.167	(0.047)	--	--	--	--	S6	5
K020201	1.075	(0.072)	1.055	(0.043)	0.084	(0.022)	--	--	--	--	S6	7
K020301	0.368	(0.042)	0.345	(0.279)	0.266	(0.060)	--	--	--	--	S6	8
K020701	1.573	(0.33[illegible])	2.798	(0.108)	0.139	(0.007)	--	--	--	--	S6	12
K020801	2.017	(0.38[illegible])	2.282	(0.061)	0.222	(0.008)	--	--	--	--	S6	13
K020901	1.265	(0.2[illegible]1)	2.478	(0.085)	0.225	(0.010)	--	--	--	--	S6	14
K021201	1.203	(0.[illegible]02)	3.038	(0.184)	0.231	(0.009)	--	--	--	--	S6	16
K021601	1.474	(0.114)	-1.895	(0.082)	0.214	(0.053)	S7	3	--	--	--	--
K021701	1.250	(0.084)	-1.297	(0.065)	0.160	(0.037)	S7	4	--	--	--	--
K022301	0.965	(0.119)	-0.230	(0.078)	0.335	(0.029)	S7	9	--	--	--	--
K022801	1.314	(0.159)	0.068	(0.040)	0.229	(0.018)	S7	14	--	--	--	--
K023201	1.033	(0.116)	-0.533	(0.081)	0.295	(0.033)	S7	17	--	--	--	--
K023501	1.951	(0.339)	0.403	(0.040)	0.222	(0.011)	S7	20	--	--	--	--
K023801	1.461	(0.080)	0.155	(0.038)	0.184	(0.024)	--	--	S7	3	S7	3
K025101	1.625	(0.134)	1.081	(0.026)	0.272	(0.012)	--	--	S7	17	S7	17
K025301	1.292	(0.100)	1.743	(0.037)	0.121	(0.008)	--	--	S7	19	S7	19
K025502	0.621	(0.026)	-0.219	(0.032)	0.000	(0.009)	--	--	S7	22	S7	22
K025503	1.045	(0.049)	2.241	(0.062)	0.000	(0.000)	--	--	S7	23	S7	23
K025801	1.319	(0.262)	2.670	(0.137)	0.100	(0.009)	--	--	--	--	S7	26
K025901	0.748	(0.018)	-0.969	(0.023)	0.000	(0.000)	S8	1	S8	1	S8	1
K026101	0.783	(0.018)	-0.044	(0.014)	0.000	(0.000)	S8	2	S8	2	S8	2
K026201	0.335	(0.013)	1.022	(0.045)	0.000	(0.000)	S8	3	S8	3	S8	3

Table E-28 (continued)
1990 IRT Parameters, Science Cross-sectional Samples
Physical Sciences Subscale

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 9/Grade 4 Block	Age 9/Grade 4 Item	Age 13/Grade 8 Block	Age 13/Grade 8 Item	Age 17/Grade 12 Block	Age 17/Grade 12 Item
K026301	0.713	(0.050)	1.395	(0.116)	0.000	(0.000)	S8	4	--	--	--	--
K026301	0.863	(0.027)	0.288	(0.016)	0.000	(0.000)	--	--	S8	4	S8	4
K026701	0.898	(0.029)	2.163	(0.049)	0.000	(0.000)	S8	8	S8	8	S8	8
K026801	0.593	(0.017)	0.426	(0.019)	0.000	(0.000)	S8	9	S8	9	S8	9
K027601	1.456	(0.086)	1.693	(0.035)	0.000	(0.000)	--	--	--	--	S8	16
K028201	1.039	(0.057)	-0.458	(0.056)	0.194	(0.026)	S9	5	S9	5	--	--
K028301	1.190	(0.095)	0.427	(0.031)	0.244	(0.014)	S9	6	S9	6	--	--
K028701	0.637	(0.039)	2.488	(0.140)	0.000	(0.000)	S9	10	S9	10	--	--
K028801	1.198	(0.037)	-0.027	(0.013)	0.000	(0.000)	S9	11	S9	11	--	--
K028802	2.630	(0.296)	1.698	(0.063)	0.000	(0.000)	S9	12	S9	12	--	--
K030201	1.326	(0.199)	2.075	(0.064)	0.255	(0.013)	--	--	--	--	S9	8
K030301	0.658	(0.073)	1.715	(0.084)	0.142	(0.029)	--	--	--	--	S9	9
N400501	0.707	(0.101)	-0.154	(0.121)	0.245	(0.044)	S3	13	--	--	--	--
N401501	0.423	(0.055)	-0.699	(0.263)	0.281	(0.061)	S3	5	--	--	--	--
N401901	0.532	(0.125)	0.770	(0.172)	0.320	(0.043)	S3	15	--	--	--	--
N402201	0.549	(0.073)	-0.148	(0.144)	0.183	(0.047)	S3	10	--	--	--	--
N403601	0.919	(0.142)	0.017	(0.079)	0.274	(0.033)	S3	14	--	--	--	--
N403701	0.716	(0.060)	-1.631	(0.157)	0.258	(0.060)	S3	4	--	--	--	--
N405101	1.154	(0.084)	0.842	(0.040)	0.242	(0.019)	--	--	S3	12	S3	12
N406701	0.899	(0.102)	0.332	(0.084)	0.196	(0.040)	--	--	S3	14	--	--
N406901	0.587	(0.036)	0.083	(0.107)	0.144	(0.040)	--	--	S3	6	S3	6
N407201	0.754	(0.059)	0.537	(0.085)	0.220	(0.035)	--	--	S3	7	S3	7
N411001	1.100	(0.150)	2.168	(0.073)	0.187	(0.014)	--	--	--	--	S3	23
N411101	0.778	(0.068)	0.724	(0.099)	0.183	(0.042)	--	--	--	--	S3	17
N411401	1.731	(0.149)	0.934	(0.035)	0.170	(0.021)	--	--	--	--	S3	16
N411801	1.808	(0.158)	0.898	(0.036)	0.174	(0.022)	--	--	--	--	S3	21

Table E-29
1990 IRT Parameters, Science Cross-sectional Samples
Earth and Space Sciences Subscale

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 9/Grade 4 Block	Age 9/Grade 4 Item	Age 13/Grade 8 Block	Age 13/Grade 8 Item	Age 17/Grade 12 Block	Age 17/Grade 12 Item
K010101	0.944	(0.045)	-0.371	(0.062)	0.205	(0.028)	S4	1	S4	1	S4	1
K010701	0.668	(0.039)	-0.067	(0.094)	0.237	(0.032)	S4	7	S4	7	S4	7
K010801	0.992	(0.054)	0.216	(0.045)	0.238	(0.019)	S4	8	S4	8	S4	8
K011001	1.576	(0.077)	0.239	(0.025)	0.220	(0.012)	S4	10	S4	10	S4	10
K011101	1.281	(0.064)	0.349	(0.029)	0.225	(0.013)	S4	11	S4	11	S4	11
K011201	1.684	(0.129)	0.931	(0.021)	0.316	(0.008)	S4	12	S4	12	S4	12
K011601	1.161	(0.077)	0.954	(0.029)	0.274	(0.011)	S4	16	S4	16	S4	16
K011801	1.382	(0.171)	1.910	(0.060)	0.276	(0.006)	S4	18	S4	18	S4	18
K013201	0.662	(0.043)	-1.029	(0.144)	0.251	(0.052)	S5	2	S5	2	--	--
K013301	1.724	(0.114)	0.559	(0.021)	0.182	(0.009)	S5	3	S5	3	--	--
K013501	2.032	(0.135)	0.092	(0.025)	0.267	(0.012)	S5	5	S5	5	--	--
K013601	0.641	(0.098)	0.968	(0.089)	0.381	(0.026)	S5	6	S5	6	--	--
K013901	1.182	(0.065)	-0.237	(0.045)	0.200	(0.021)	S5	10	S5	10	--	--
K014701	1.203	(0.086)	0.545	(0.033)	0.224	(0.013)	S5	17	S5	17	--	--
K014901	1.907	(0.495)	1.796	(0.134)	0.245	(0.006)	S5	19	S5	19	--	--
K015401	1.073	(0.110)	0.360	(0.077)	0.256	(0.036)	--	--	S5	23	--	--
K015402	0.974	(0.137)	1.202	(0.058)	0.187	(0.024)	--	--	S5	24	--	--
K015501	2.187	(0.225)	0.586	(0.029)	0.221	(0.017)	--	--	S5	25	--	--
K016301	1.751	(0.328)	1.625	(0.050)	0.438	(0.015)	--	--	--	--	S5	7
K016701	1.003	(0.098)	0.570	(0.096)	0.243	(0.044)	--	--	--	--	S5	12
K016802	1.199	(0.142)	1.214	(0.060)	0.308	(0.025)	--	--	--	--	S5	14
K017301	1.745	(0.212)	1.491	(0.037)	0.247	(0.015)	--	--	--	--	S5	19
K017401	1.434	(0.153)	1.561	(0.037)	0.135	(0.015)	--	--	--	--	S5	20
K018201	0.966	(0.069)	-0.476	(0.079)	0.249	(0.034)	S6	4	S6	4	--	--
K018701	0.870	(0.027)	-0.509	(0.019)	0.000	(0.000)	S6	10	S6	10	--	--
K019401	1.128	(0.127)	0.584	(0.061)	0.218	(0.030)	--	--	S6	18	--	--
K019801	0.381	(0.077)	1.803	(0.286)	0.341	(0.055)	--	--	--	--	S6	4
K020601	1.349	(0.176)	1.250	(0.051)	0.292	(0.024)	--	--	--	--	S6	11
K021801	1.230	(0.099)	-1.416	(0.095)	0.227	(0.053)	S7	5	--	--	--	--
K021901	1.402	(0.159)	-0.651	(0.064)	0.224	(0.033)	S7	6	--	--	--	--
K022401	1.186	(0.406)	1.110	(0.268)	0.329	(0.014)	S7	10	--	--	--	--
K022501	1.500	(0.176)	-0.076	(0.034)	0.148	(0.018)	S7	11	--	--	--	--
K022601	0.958	(0.144)	0.018	(0.064)	0.217	(0.030)	S7	12	--	--	--	--
K023301	1.351	(0.250)	0.554	(0.080)	0.186	(0.014)	S7	18	--	--	--	--
K023401	1.535	(0.311)	0.497	(0.075)	0.257	(0.015)	S7	19	--	--	--	--
K024201	0.682	(0.074)	0.869	(0.115)	0.311	(0.038)	--	--	S7	8	S7	8
K024301	1.367	(0.080)	0.850	(0.027)	0.119	(0.014)	--	--	S7	9	S7	9
K024501	1.405	(0.137)	1.347	(0.033)	0.312	(0.012)	--	--	S7	11	S7	11
K024502	2.122	(0.265)	1.564	(0.031)	0.364	(0.008)	--	--	S7	12	S7	12
K024701	1.550	(0.111)	1.150	(0.024)	0.196	(0.011)	--	--	S7	14	S7	14
K024801	0.899	(0.071)	1.045	(0.050)	0.181	(0.022)	--	--	S7	15	S7	15
K025201	1.242	(0.092)	1.356	(0.028)	0.134	(0.012)	--	--	S7	18	S7	18
K025401	1.873	(0.533)	2.331	(0.134)	0.264	(0.006)	--	--	S7	20	S7	20
K026401	1.011	(0.025)	0.955	(0.018)	0.000	(0.000)	S8	5	S8	5	S8	5
K026501	1.218	(0.028)	0.165	(0.011)	0.000	(0.000)	S8	6	S8	6	S8	6
K027101	0.723	(0.027)	-0.405	(0.031)	0.000	(0.000)	--	--	S8	11	S8	11
K027301	0.400	(0.031)	3.351	(0.206)	0.000	(0.000)	--	--	S8	13	S8	13
K027501	1.133	(0.076)	1.956	(0.054)	0.000	(0.000)	--	--	--	--	S8	15
K027701	0.576	(0.034)	-1.452	(0.158)	0.229	(0.055)	S9	1	S9	1	--	--
K028401	0.798	(0.087)	0.864	(0.054)	0.233	(0.020)	S9	7	S9	7	--	--
K028601	1.657	(0.140)	0.480	(0.029)	0.293	(0.012)	S9	9	S9	9	--	--
K028901	0.496	(0.053)	-0.091	(0.200)	0.247	(0.057)	--	--	S9	13	--	--
K030101	1.135	(0.107)	0.710	(0.073)	0.212	(0.036)	--	--	--	--	S9	7

Table E-29 (continued)
1990 IRT Parameters, Science Cross-sectional Samples
Earth and Space Sciences Subscale

NAEP ID	A	S.E.	B	S.E.	C	S.E.	Age 9/Grade 4 Block	Age 9/Grade 4 Item	Age 13/Grade 8 Block	Age 13/Grade 8 Item	Age 17/Grade 12 Block	Age 17/Grade 12 Item
K030602	0.546	(0.039)	1.784	(0.072)	0.000	(0.000)	--	--	--	--	S9	11
N401601	1.106	(0.125)	-1.009	(0.108)	0.220	(0.053)	S3	6	--	--	--	--
N402101	0.967	(0.114)	-0.744	(0.103)	0.202	(0.047)	S3	7	--	--	--	--
N402401	0.585	(0.174)	1.299	(0.305)	0.263	(0.034)	S3	17	--	--	--	--
N402501	2.019	(0.497)	0.440	(0.066)	0.283	(0.014)	S3	16	--	--	--	--
N405501	1.080	(0.092)	0.522	(0.070)	0.288	(0.031)	--	--	S3	9	S3	9
N406501	0.960	(0.108)	0.746	(0.063)	0.154	(0.031)	--	--	S3	16	--	--
N406601	0.515	(0.042)	-0.234	(0.191)	0.231	(0.058)	--	--	S3	3	S3	3
N407301	0.462	(0.049)	1.099	(0.143)	0.183	(0.041)	--	--	S3	10	S3	10
N407302	1.343	(0.198)	1.665	(0.051)	0.395	(0.013)	--	--	S3	11	S3	11
N408101	1.049	(0.181)	1.961	(0.074)	0.181	(0.021)	--	--	--	--	S3	25
N410201	0.825	(0.170)	2.099	(0.109)	0.233	(0.026)	--	--	--	--	S3	24

Table E-30
1990 IRT Parameters, Science Cross-sectional Samples
Nature of Science Subscale

							Age 9/Grade 4		Age 13/Grade 8		Age 17/Grade 12	
NAEP ID	A	S.E.	B	S.E.	C	S.E.	Block	Item	Block	Item	Block	Item
K010301	1.196	(0.086)	-0.563	(0.091)	0.261	(0.046)	S4	3	S4	3	S4	3
K011301	1.239	(0.072)	-0.321	(0.061)	0.168	(0.032)	S4	13	S4	13	S4	13
K011901	1.089	(0.430)	1.343	(0.338)	0.249	(0.019)	S4	19	--	--	--	--
K011901	0.979	(0.075)	0.205	(0.077)	0.185	(0.035)	--	--	S4	19	S4	19
K012601	0.761	(0.125)	2.061	(0.109)	0.139	(0.018)	--	--	S4	26	S4	26
K012901	0.919	(0.085)	0.395	(0.086)	0.129	(0.038)	--	--	--	--	S4	29
K014501	1.833	(0.232)	0.027	(0.048)	0.224	(0.024)	S5	15	S5	15	--	--
K014601	1.417	(0.176)	0.288	(0.049)	0.225	(0.023)	S5	16	S5	16	--	--
K016001	1.773	(0.231)	0.686	(0.058)	0.167	(0.032)	--	--	--	--	S5	4
K016401	1.326	(0.204)	0.993	(0.065)	0.212	(0.033)	--	--	--	--	S5	8
K018101	0.829	(0.092)	-0.269	(0.139)	0.280	(0.052)	S6	3	S6	3	--	--
K018301	1.238	(0.129)	-0.267	(0.084)	0.224	(0.041)	S6	5	S6	5	--	--
K019201	0.851	(0.234)	1.648	(0.174)	0.223	(0.028)	--	--	S6	16	--	--
K022701	1.294	(0.281)	-0.041	(0.089)	0.212	(0.041)	S7	13	--	--	--	--
K023101	1.358	(0.530)	0.952	(0.209)	0.219	(0.019)	S7	16	--	--	--	--
K023601	1.508	(0.136)	0.401	(0.053)	0.192	(0.029)	--	--	S7	1	S7	1
K024101	0.965	(0.106)	0.749	(0.079)	0.263	(0.033)	--	--	S7	7	S7	7
K024601	1.359	(0.185)	1.429	(0.038)	0.193	(0.015)	--	--	S7	13	S7	13
K025601	0.555	(0.094)	1.261	(0.164)	0.196	(0.050)	--	--	--	--	S7	24
K028101	1.804	(0.222)	-0.283	(0.068)	0.275	(0.036)	S9	4	S9	4	--	--
K029401	0.870	(0.186)	1.048	(0.090)	0.236	(0.036)	--	--	S9	17	--	--
K029601	1.298	(0.163)	0.992	(0.053)	0.136	(0.028)	--	--	--	--	S9	2
K029701	1.625	(0.259)	0.977	(0.053)	0.208	(0.029)	--	--	--	--	S9	3
N403001	1.039	(0.218)	-2.487	(0.243)	0.176	(0.054)	S3	3	--	--	--	--
N404001	0.537	(0.096)	-0.137	(0.181)	0.173	(0.052)	S3	9	--	--	--	--
N409001	0.980	(0.118)	0.244	(0.089)	0.158	(0.043)	--	--	S3	15	--	--
N409201	1.482	(0.423)	1.369	(0.103)	0.317	(0.020)	--	--	S3	17	--	--
N409501	1.370	(0.181)	1.409	(0.035)	0.131	(0.015)	--	--	S3	13	S3	13
N409601	1.161	(0.374)	1.728	(0.193)	0.342	(0.021)	--	--	S3	20	--	--
N410101	1.081	(0.140)	-0.043	(0.171)	0.431	(0.063)	--	--	--	--	S3	18
N410102	0.851	(0.129)	0.323	(0.197)	0.427	(0.060)	--	--	--	--	S3	19
N410103	0.979	(0.129)	-0.069	(0.187)	0.439	(0.064)	--	--	--	--	S3	20

APPENDIX F

Conditioning Variables and Related Data

Appendix F

CONDITIONING VARIABLES AND RELATED DATA

This appendix contains information about the conditioning variables used in the construction of plausible values for the 1990 assessments of reading, mathematics, and science.

The first part of the appendix gives the contrast codings for each set of conditioning variables used in 1990. Codings for the common conditioning variables, which were used for cross-sectional studies in reading, mathematics, and science, are given in Table F-1. In addition, subject-specific conditioning variables used in each cross-sectional and trend study are given for reading (Tables F-2 and F-3), mathematics (Tables F-4 and F-5), and science (Tables F-6 and F-7).

Some conditioning variables were constructed by recoding the values of a data variable or by combining and recoding data from two or more variables. A description of how these conditioning variables were derived is included in Appendix B.

The second part of the appendix (beginning with Table F-8) shows the estimated effects from the conditioning model for the variables defined by the codings in Tables F-1 through F-7 or for principal components of those variables (see Chapter 11 for details of conditioning). Principal components of the contrast codings of the conditioning variables were used in the mathematics and science cross-sectional analyses to eliminate multicollinearity. This part of the appendix also includes the proportion of the variance in the conditioning variable contrasts that is accounted for by the principal components used in the conditioning model for mathematics (Tables F-15, F-17, and F-19) and for science (Tables F-24, F-26, and F-28).

Note that all effect estimates are in the metrics used in the original calibration of the scale. The transformations needed to represent these effects in terms of the metric of the final reporting scales appear in the chapters that describe the scaling of each subject area. Note also that certain conditioning variables in the tables for trend and reading cross-sectional analyses do not have effect estimates. This is because those variables are approximate linear combinations of the other conditioning variables, and so effects for those variables were not calculated.

Table F-1

Contrast Codings for 1990 NAEP Common Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Overall	All	---	---	1
Gender	All	DSEX	1 Male 2 Female	0 1
Ethnicity	All	DRACE	1 White 2 Black 3 Hispanic 4 Asian American 5 American Indian 6 Unclassified Missing	000 100 010 001 000 000 000
Size and Type of Community	All	STOC	2 Low Metro 3 High Metro 1, 4-7 All others and Missing	00 10 01
Region	All	REGION	1 Northeast 2 Southeast 3 Central 4 West	000 100 010 001
Parents' Education	All	PARED	1 Less than high school 2 High school graduate 3 Post-high school 4 College graduate Missing and I Don't Know	0000 1000 0100 0010 0001
Items in the Home (of newspaper, > 25 books, encyclopedia, magazines. Two or more missing = Missing.)	All	HOMEEN2	1 0 to 2 items 2 3 items 3 4 items Missing	00 10 01 00
TV Watching	All	B001801	1 None 2 One hour or less per day 3 Two hours per day 4 Three hours per day 5 Four hours per day 6 Five hours per day 7 Six or more hours per day Missing	0 $\overline{00}$ 1 01 2 04 3 09 4 16 5 25 6 36 3 09

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-1 (continued)

Contrast Codings for 1990 NAEP Common Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Home Language Minority (How often do people in your home speak a language other than English?)	All	B003201	1 Never 2, 3 Sometimes, Always Missing	0 1 0
Homework	9	B006601	1 Don't have any 2 Don't do any 3 ½ hour 4 One hour 5 More than one hour Missing	1000 0100 0111 0124 0139 0000
	13, 17	B003901	1 Don't have any 2 Don't do any 3 ½ hour 4 One hour 5 Two hours 6 More than two hours Missing	100 $\overline{00}$ 010 00 011 01 012 04 013 09 014 16 000 00
Percent in Lunch Program	All	PCLUNCH	0 1 2 . . 99 100 Missing	$\overline{000}$ 0 001 0 002 0 . . 099 0 100 0 000 1
Percent White in School	All	PCTWHT	0 -49 White minority 50-79 Integrated 80-100 Predominantly White Missing	10 01 00 00
Age by Grade (category 4 not applicable for age class 17)	All	MODAGE MODGRD	1 < modal age, modal grade 2 Modal age, < modal grade 3 Modal age, modal grade, and Missing 4 Modal age, > modal grade 5 > modal age, modal grade	0000 1000 0100 0010 0001

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-1 (continued)

Contrast Codings for 1990 NAEP Common Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
School Type	All	SCHTYPE	1 Public 2 Private 3 Catholic 4 Bureau of Indian Affairs 5 Department of Defense Missing	0 1 1 1 1 1
Someone at Home Helps with Homework	All	B006701	1 Almost every day 2 Once or twice a week 3 Once or twice a month 4 Never 5 Don't have homework Missing	1 1 0 0 0 0
Single/multiple Parent(s) at Home	All	SINGLEP	1 Yes to father and mother 2 Any other responses Missing	1 0 0
Mother at Home	All	B005601	1 Yes 2 No Missing	1 0 0
Mother Works Outside of Home	9	B005901	1 Yes 2 No 3 Mother not at home Missing	1 0 0 0
	13, 17	B006001	1 Yes, full-time 2 Yes, part-time 3 No 4 Mother not at home Missing	1 1 0 0 0
Pages Read for School and Homework Each Day	All	B001101	1 More than 20 pages 2 16 - 20 pages 3 11 - 15 pages 4 6 - 10 pages 5 5 or fewer pages Missing	11 11 11 10 00 00
Season of Assessment	All	SEASON	Winter Spring	1 0

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-1 (continued)

Contrast Codings for 1990 NAEP Common Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
School Level Average Mathematics Proficiency	All	SCHMATH	Not missing Missing	1 [mean] 0 0
School Level Average Reading Proficiency	All	SCHREAD	Not missing Missing	1 [mean] 0 0
School Level Average Science Proficiency	All	SCHSCI	Not missing Missing	1 [mean] 0 0
Went to Preschool	9	B004201	1 Yes 2 No 3 I Don't Know Missing	1 0 0 0
Do You Expect to Graduate from High School	13	S003401	1 Yes 2 No Missing	1 0 0
Days of School Missed Last Month	13, 17	S004001	1 None 2 1 or 2 days 3 3 or 4 days 4 5 to 10 days 5 More than 10 days Missing	1 1 0 0 0 0
Rules of Behavior Are Strict	13, 17	B007001	1 Strongly agree 2 Agree 3 Disagree 4 Strongly disagree Missing	11 12 13 14 00
Don't Feel Safe at This School	13, 17	B007002	1 Strongly agree 2 Agree 3 Disagree 4 Strongly disagree Missing	11 12 13 14 00
Students Often Disrupt Class	13, 17	B007003	1 Strongly agree 2 Agree 3 Disagree 4 Strongly disagree Missing	11 12 13 14 00

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-1 (continued)

Contrast Codings for 1990 NAEP Common Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
High School Program	17	B005001	1 General 2 College preparatory 3 Vocational, technical Missing	00 10 01 00
Post-secondary Plans	17	B005501	1 Work full time 2 Two-year college 3 Four-year college 4 Other Missing	00 10 01 00 00
Number of Semesters of English/Literature/Writing	17	B007101	1 None 2 1 3 2 4 3 5 4 6 5 7 6 8 7 9 8 or more Missing	10 11 12 13 14 15 16 17 18 00
Number of Semesters of Mathematics	17	B007102	1 None 2 1 3 2 4 3 5 4 6 5 7 6 8 7 9 8 or more Missing	10 11 12 13 14 15 16 17 18 00

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-1 (continued)

Contrast Codings for 1990 NAEP Common Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Number of Semesters of Science	17	B007103	1 None	10
			2 1	11
			3 2	12
			4 3	13
			5 4	14
			6 5	15
			7 6	16
			8 7	17
			9 8 or more	18
			Missing	00
Number of Semesters of History/Social Studies/Geography	17	B007104	1 None	10
			2 1	11
			3 2	12
			4 3	13
			5 4	14
			6 5	15
			7 6	16
			8 7	17
			9 8 or more	18
			Missing	00
Number of Semesters of Foreign Language(s)	17	B007105	1 None	10
			2 1	11
			3 2	12
			4 3	13
			5 4	14
			6 5	15
			7 6	16
			8 7	17
			9 8 or more	18
			Missing	00

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-1 (continued)

Contrast Codings for 1990 NAEP Common Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Number of Semesters of Vocational/Technical/ Business Education	17	B007106	1 None	10
			2 1	11
			3 2	12
			4 3	13
			5 4	14
			6 5	15
			7 6	16
			8 7	17
			9 8 or more	18
			Missing	00
Number of Semesters of Art/Music	17	B007107	1 None	10
			2 1	11
			3 2	12
			4 3	13
			5 4	14
			6 5	15
			7 6	16
			8 7	17
			9 8 or more	18
			Missing	00

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-2

Contrast Codings for 1990 NAEP Reading Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Read Outside of School	All	R810001	1 Almost daily 2 1-2 times a week 3 1-2 times a month 4 Few times a year 5 Never Other	1 2 3 4 5 5
Tell Others About Book	All	R81002	1 Almost daily 2 1-2 times a week 3 1-2 times a month 4 Few times a year 5 Never Other	1 $\overline{01}$ 2 04 3 09 4 16 5 25 5 25
Take Out Library Book for Self	All	R810003	1 Almost daily 2 1-2 times a week 3 1-2 times a month 4 Few times a year 5 Never Other	1 $\overline{01}$ 2 04 3 09 4 16 5 25 5 25
What Kind of Reader Are You	All	R810201	1 Very good 2 Good 3 Average 4 Poor Other	1 2 3 4 4
Teacher Asks About Book	9	R810102	1 Almost daily 2 1-2 times a week 3 1-2 times a month 4 Few times a year 5 Never Other	1 $\overline{01}$ 2 04 3 09 4 16 5 25 5 25
Asked to Work in Workbook	9	R810103	1 Almost daily 2 1-2 times a week 3 1-2 times a month 4 Few times a year 5 Never Other	1 $\overline{01}$ 2 04 3 09 4 16 5 25 5 25

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-2 (continued)

Contrast Codings for 1990 NAEP Reading Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Asked to Write About Book	9	R810104	1 Almost daily 2 1-2 times a week 3 1-2 times a month 4 Few times a year 5 Never Other	1 $\overline{01}$ 2 04 3 09 4 16 5 25 1 01
Asked to Do a Group Activity	9	R810105	1 Almost daily 2 1-2 times a week 3 1-2 times a month 4 Few times a year 5 Never Other	1 $\overline{01}$ 2 04 3 09 4 16 5 25 1 01
Teacher Teaches New Words	9	R810101	1 Almost daily 2 1-2 times a week 3 1-2 times a month 4 Few times a year 5 Never Other	1 $\overline{01}$ 2 04 3 09 4 16 5 25 5 25
Read Out Loud	9 [Rdg-ABB]	R800501	1 Almost daily 2 1-2 times a week 3 1-2 times a month 4 Never/hardly ever Other	1 $\overline{01}$ 2 04 3 09 4 16 1 01
Read for Fun	9 [Rdg-ABB]	S003501	1 Almost daily 2 1-2 times a week 3 1-2 times a month 4 Few times a year 5 Never/hardly ever Other	1 2 3 4 5 5
Tell a Friend About a Book	9 [Rdg-ABB]	S003502	1 Almost daily 2 1-2 times a week 3 1-2 times a month 4 Few times a year 5 Never/hardly ever Other	1 $\overline{01}$ 2 04 3 09 4 16 5 25 1 01

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-2 (continued)

Contrast Codings for 1990 NAEP Reading Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Use Library	9 [Rdg-ABB]	S003503	1 Almost daily 2 1-2 times a week 3 1-2 times a month 4 Few times a year 5 Never/hardly ever Other	1 $\overline{01}$ 2 04 3 09 4 16 5 25 5 25
Spend Your Money on Book	9 [Rdg-ABB]	S003504	1 Almost daily 2 1-2 times a week 3 1-2 times a month 4 Few times a year 5 Never/hardly ever Other	1 $\overline{01}$ 2 04 3 09 4 16 5 25 1 01
What Kind of Reader Are You	9 [Rdg-ABB]	S003301	1 Poor 2 Good 3 Very good 4 I don't know Other	1 2 3 1 1
Asked to Choose Book	9 [Rdg-ABB]	R800701	1 Almost daily 2 1-2 times a week 3 1-2 times a month 4 Never/hardly ever Other	1 2 3 4 4
Get Any Magazines for Yourself	9 [Rdg-ABB]	R800101	1 Yes 2 No Other	1 2 2
Own Any Nonschool Books	9 [Rdg-ABB]	R800301	1 10 or fewer 2 11 - 20 3 21 - 30 4 31 or more Other	1 2 3 4 1
Are You Read to at Home	9 [Rdg-ABB]	R800401	1 Almost daily 2 1-2 times a week 3 1-2 times a month 4 Hardly ever 5 I don't know Other	1 2 3 4 4 4

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-2 (continued)

Contrast Codings for 1990 NAEP Reading Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Read More by Same Author	13, 17	R810301	1 Weekly 2 Monthly 3 Few times a year 4 Never Other	1 2 3 4 4
Teacher Discusses New Vocabulary	13, 17	R810401	1 Almost daily 2 1-2 times a week 3 1-2 times a month 4 Few times a year 5 Never Other	1 $\overline{01}$ 2 04 3 09 4 16 5 25 5 25
Teacher Asks About Book	13, 17	R810402	1 Almost daily 2 1-2 times a week 3 1-2 times a month 4 Few times a year 5 Never Other	1 2 3 4 5 5
Asked to Work in Workbook	13, 17	R810403	1 Almost daily 2 1-2 times a week 3 1-2 times a month 4 Few times a year 5 Never Other	1 $\overline{01}$ 2 04 3 09 4 16 5 25 5 25
Asked to Write in Journal	13, 17	R810404	1 Almost daily 2 1-2 times a week 3 1-2 times a month 4 Few times a year 5 Never Other	1 2 3 4 5 1
Asked to Write Report About Book	13, 17	R810405	1 Almost daily 2 1-2 times a week 3 1-2 times a month 4 Few times a year 5 Never Other	1 $\overline{01}$ 2 04 3 09 4 16 5 25 1 01

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-2 (continued)

Contrast Codings for 1990 NAEP Reading Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Asked to Do Activity About Book	13, 17	R810406	1 Almost daily 2 1-2 times a week 3 1-2 times a month 4 Few times a year 5 Never Other	1 $\overline{01}$ 2 04 3 09 4 16 5 25 1 01
Asked to Think About Author	13, 17	R810407	1 Almost daily 2 1-2 times a week 3 1-2 times a month 4 Few times a year 5 Never Other	1 $\overline{01}$ 2 04 3 09 4 16 5 25 1 01
Asked to Explain Your Understanding	13, 17	R810408	1 Almost daily 2 1-2 times a week 3 1-2 times a month 4 Few times a year 5 Never Other	1 2 3 4 5 5
Asked to Discuss Different Interpretations	13, 17	R810409	1 Almost daily 2 1-2 times a week 3 1-2 times a month 4 Few times a year 5 Never Other	1 2 3 4 5 5
Asked to Predict What You Find	13, 17	R810410	1 Almost daily 2 1-2 times a week 3 1-2 times a month 4 Few times a year 5 Never Other	1 $\overline{01}$ 2 04 3 09 4 16 5 25 1 01
What Kind of Reading in Spare Time	13, 17	R810501	1 Usually don't 2 Fiction 3 Nonfiction 4 Both fiction and nonfiction Other	1 3 2 4 1

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-2 (continued)

Contrast Codings for 1990 NAEP Reading Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
How Much Reading Adults in Home Do	13, 17	R810601	1 A lot 2 Some 3 Hardly at all 4 None Other	1 2 3 4 4
Teacher Shows How to Use Reading Skills	17	R810413	1 Almost daily 2 1-2 times a week 3 1-2 times a month 4 Few times a year 5 Never Other	1 $\overline{01}$ 2 04 3 09 4 16 5 25 1 01
Read Textbook for Assignment	17	R810701	1 Almost daily 2 1-2 times a week 3 1-2 times a month 4 Few times a year 5 Never Other	1 2 3 4 5 5
Read Newspaper/Magazine for Assignment	17	R810702	1 Almost daily 2 1-2 times a week 3 1-2 times a month 4 Few times a year 5 Never Other	1 $\overline{01}$ 2 04 3 09 4 16 5 25 1 01
Read Story/Poem for Assignment	17	R810703	1 Almost daily 2 1-2 times a week 3 1-2 times a month 4 Few times a year 5 Never Other	1 2 3 4 5 5
Read Encyclopedia/Dictionary for Assignment	17	R810704	1 Almost daily 2 1-2 times a week 3 1-2 times a month 4 Few times a year 5 Never Other	1 $\overline{01}$ 2 04 3 09 4 16 5 25 5 25

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-3

Contrast Codings for 1990 NAEP Reading Trend Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Overall	All		---	1
Gender	All	DSEX	1 Male 2 Female	0 1
Size and Type of Community	All	STOC	1 Low Metro 2 High Metro 3 All others and Missing	00 10 01
Region	All	REGION	1 Northeast 2 Southeast 3 Central 4 West	000 100 010 001
Parents' Education	All	PARED	1 Less than high school 2 High school graduate 3 Post-high school 4 College graduate Missing and I Don't Know	0000 1000 0100 0010 0001
Items in the Home	All	B000901 B000902 B000903 B000904 B000905 B000906	0 None of the six items 1 One of the six items 2 Two of the six items 3 Three of the six items 4 Four of the six items 5 Five of the six items 6 Six of the six items Missing	00 10 20 30 40 50 60 01
Television Watching	All	B001801	1 None 2 One hour or less 3 Two hours 4 Three hours 5 Four hours 6 Five hours 7 Six or more hours Missing	00 10 20 30 40 50 60 01

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-3 (continued)

Contrast Codings for 1990 NAEP Reading Trend Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Homework	All	B001701	1 Don't have any 2 Don't do any 3 Less than 1 hour 4 1-2 hours 5 More than 2 hours Missing	00 00 10 20 30 01
Language Spoken at Home	All	B000401	1 English 2 Spanish 3 Other Missing	00 10 10 01
Pages Read	All	B001101	1 More than 20 2 16-20 3 11-15 4 6-10 5 5 or fewer Missing	10 10 10 10 00 01
Percent in School Lunch Program	All	PCLUNCH	0 0 percent 1 1 percent 2 2 percent . . 99 99 percent 100 100 percent Missing	$\overline{000}$ 0 001 0 002 0 . . 099 0 100 0 000 1
Percent White	All	PCTWHT	0-49 White minority 50-79 Integrated 80-100 Predominantly White Missing	100 010 001 000
Courses Taken	9, 13	B001001 B001002 B002003 B002004 B002005 B002006 B002007	0 None of the seven 1 One of the seven 2 Two of the seven 3 Three of the seven 4 Four of the seven 5 Five of the seven 6 Six of the seven 7 Seven of the seven Missing	$\overline{00.0}$ 0 01.0 0 02.0 0 03.0 0 04.0 0 05.0 0 06.0 0 07.0 0 00.0 1

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-3 (continued)

Contrast Codings for 1990 NAEP Reading Trend Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Derived Race/Ethnicity	All	DRACE	1 White 2 Black 3 Hispanic 4 Asian American 5 American Indian 6 Unclassified Missing	000 100 010 001 000 000 000
Age by Grade	All	MODGRD MODAGE	1 < modal age, modal grade 2 Modal age, < modal grade 3 Modal age, modal grade/missing 4 Modal age, > modal grade 5 > modal age, modal grade	0000 1000 0100 0010 0001

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-4

Contrast Codings for 1990 NAEP Mathematics Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Do Textbook Problems	All	M810101	1 Almost every day 2 Several times a week 3 Once a week 4 Less than once a week 5 Never Missing	1000 0100 0010 0001 0000 0000
Do Worksheet Problems	All	M810102	1 Almost every day 2 Several times a week 3 Once a week 4 Less than once a week 5 Never Missing	1000 0100 0010 0001 0000 1000
Work in Small Groups	All	M810103	1 Almost every day 2 Several times a week 3 Once a week 4 Less than once a week 5 Never Missing	1000 0100 0010 0001 0000 1000
Work with Objects	All	M810112	1 Almost every day 2 Several times a week 3 Once a week 4 Less than once a week 5 Never Missing	1000 0100 0010 0001 0000 1000
Uses Calculator	All	M810105	1 Almost every day 2 Several times a week 3 Once a week 4 Less than once a week 5 Never Missing	1000 0100 0010 0001 0000 0000
Uses Computer	All	M810106	1 Almost every day 2 Several times a week 3 Once a week 4 Less than once a week 5 Never Missing	1000 0100 0010 0001 0000 1000

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-4 (continued)

Contrast Codings for 1990 NAEP Mathematics Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Takes Tests	All	M810107	1 Almost every day 2 Several times a week 3 Once a week 4 Less than once a week 5 Never Missing	1000 0100 0010 0001 0000 1000
Teacher Explains Calculator Use	All	M810201	1 Yes 2 No Missing	10 01 00
Amount of Homework Done	All	M810601	1 None 2 15 minutes 3 30 minutes 4 45 minutes 5 60 minutes 6 More than an hour 7 Not taking math Missing	1000 0100 0100 0010 0001 0001 0000 0000
Likes Mathematics	9	M811101	1 Agree 2 Undecided 3 Disagree Missing	01 10 00 00
People Use Math	9	M811102	1 Agree 2 Undecided 3 Disagree Missing	01 10 00 00
Good at Math	9	M811103	1 Agree 2 Undecided 3 Disagree Missing	01 10 00 00
Math is for Boys	9	M811104	1 Agree 2 Undecided 3 Disagree Missing	10 01 00 00

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-4 (continued)

Contrast Codings for 1990 NAEP Mathematics Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Useful for Everyday Problems	9	M8111051	1 Agree 2 Undecided 3 Disagree Missing	01 10 00 00
What Kind of Class Are You Taking	13	M810501	1 Not taking math 2 Eighth grade math 3 Pre-algebra 4 Algebra 5 Other Missing	0000 1000 0100 0010 0001 0000
Writes Reports	13, 17	M810108	1 Almost every day 2 Several times a week 3 Once a week 4 Less than once a week 5 Never Missing	1000 0100 0010 0001 0000 1000
Uses Calculator on Class Problems	13, 17	M810301	1 Almost always 2 Sometimes 3 Never Missing	10 01 00 10
Uses Calculator on Home Problems	13, 17	M810302	1 Almost always 2 Sometimes 3 Never Missing	10 01 00 10
Uses Calculator on Tests	13, 17	M810303	1 Almost always 2 Sometimes 3 Never Missing	10 01 00 10
Likes Mathematics	13, 17	M810701	1 Strongly agree 2 Agree 3 Undecided 4 Disagree 5 Strongly disagree Missing	0001 0010 0100 1000 0000 0000

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-4 (continued)

Contrast Codings for 1990 NAEP Mathematics Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
People Use Math	13, 17	M810702	1 Strongly agree 2 Agree 3 Undecided 4 Disagree 5 Strongly disagree Missing	0001 0010 0100 1000 0000 0000
Good at Math	13, 17	M810703	1 Strongly agree 2 Agree 3 Undecided 4 Disagree 5 Strongly disagree Missing	0001 0010 0100 1000 0000 0000
Math is for Boys	13, 17	M810704	1 Strongly agree 2 Agree 3 Undecided 4 Disagree 5 Strongly disagree Missing	100 100 010 010 001 000
Useful for Everyday Problems	13, 17	M810705	1 Strongly agree 2 Agree 3 Undecided 4 Disagree 5 Strongly disagree Missing	0001 0010 0100 1000 0000 0000
Writes Proofs	17	M810109	1 Almost every day 2 Several times a week 3 Once a week 4 Less than once a week 5 Never Missing	1000 0100 0010 0001 0000 1000
Formulates Problems	17	M810110	1 Almost every day 2 Several times a week 3 Once a week 4 Less than once a week 5 Never Missing	1000 0100 0010 0001 0000 1000

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-4 (continued)

Contrast Codings for 1990 NAEP Mathematics Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Are You Taking Math Classes	17	M810901	1 Yes 2 No 3 Omit 4 Not reached	10 01 00 00
How Long General Math	17	M811001	1 More than a year 2 1 year 3 ½ year 4 Not studied Missing	100 010 010 001 000
How Long Consumer Math	17	M811002	1 More than a year 2 1 year 3 ½ year 4 Not studied Missing	1000 0100 0010 0001 0000
How Long Pre-algebra	17	M811003	1 More than a year 2 1 year 3 ½ year 4 Not studied Missing	1000 0100 0010 0001 0000
How Long Algebra	17	M811004	1 More than a year 2 1 year 3 ½ year 4 Not studied Missing	1000 0100 0010 0001 0000
How Long Geometry	17	M811005	1 More than a year 2 1 year 3 ½ year 4 Not studied Missing	1000 0100 0010 0001 0000
How Long 2nd Algebra	17	M811006	1 More than a year 2 1 year 3 ½ year 4 Not studied Missing	1000 0100 0010 0001 0000

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-4 (continued)

Contrast Codings for 1990 NAEP Mathematics Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
How Long Trigonometry	17	M811007	1 More than a year 2 1 year 3 ½ year 4 Not studied Missing	1000 0100 0010 0001 0000
How Long Pre-calculus	17	M811008	1 More than a year 2 1 year 3 ½ year 4 Not studied Missing	1000 0100 0010 0001 0000
How Long Prob/Stat	17	M811009	1 More than a year 2 1 year 3 ½ year 4 Not studied Missing	1000 0100 0010 0001 0000
How Long Analytical Geometry	17	M811010	1 More than a year 2 1 year 3 ½ year 4 Not studied Missing	1000 0100 0010 0001 0000
How Long Calculus	17	M811011	1 More than a year 2 1 year 3 ½ year 4 Not studied Missing	1000 0100 0010 0001 0000
Geometry Courses Taken (Teacher)	9, 13	T030408	1 None 2 One 3 Two 4 Three Omit Unmatched	100 010 001 001 100 000
Abstract Algebra Courses Taken (Teacher)	9, 13	T030410	1 None 2 One 3 Two 4 Three Omit Unmatched	100 010 001 001 100 000

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-4 (continued)

Contrast Codings for 1990 NAEP Mathematics Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Calculus Courses Taken (Teacher)	9, 13	T030411	1 None	100
			2 One	010
			3 Two	010
			4 Three	001
			Omit	100
			Unmatched	000
Amount of Resources (Teacher)	9, 13	T030801	1 All I need	100
			2 Most of what I need	010
			3 Some of what I need	001
			4 No resources	001
			Omit	001
			Unmatched	000
Ability of Class (Teacher)	9, 13	T031001	1 Primarily high	100
			2 Primarily average	010
			3 Primarily low	001
			4 Mixed	010
			Omit	010
			Unmatched	000
Amount of Math Homework Done (Teacher)	9, 13	T031201	1 None	11
			2 15 minutes	11
			3 30 minutes	21
			4 45 minutes	31
			5 60 minutes	31
			6 More than an hour	21
			Omit	21
			Unmatched	00
Do Textbook Problems (Teacher)	9, 13	T031401	1 Almost every day	100
			2 Several times a week	010
			3 Once a week	001
			4 Less than once a week	001
			5 Never	001
			Missing	001
				000

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-4 (continued)

Contrast Codings for 1990 NAEP Mathematics Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Do Worksheet Problems (Teacher)	9, 13	T031402	1 Almost every day 2 Several times a week 3 Once a week 4 Less than once a week 5 Never Missing	100 100 010 001 001 100 000
Emphasis on Measurement Topic (Teacher)	9, 13	T031504	1 Heavy 2 Moderate 3 Little 4 None Omit Unmatched	31 21 11 01 00 00
Emphasis on Algebra Topic (Teacher)	9, 13	T031508	1 Heavy 2 Moderate 3 Little 4 None Omit Unmatched	31 21 11 01 00 00
Work in Small Groups (Teacher)	9	T031403	1 Almost every day 2 Several times a week 3 Once a week 4 Less than once a week 5 Never Missing	100 100 010 001 001 000 000
Emphasis on Communicating Ideas (Teacher)	9	T031512	1 Heavy 2 Moderate 3 Little 4 None Omit Unmatched	31 21 11 01 00 00
Use Calculators in Class (Teachers)	9, 13	T031601	1 Yes 2 No Omit Unmatched	11 01 01 00

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-4 (continued)

Contrast Codings for 1990 NAEP Mathematics Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Use Calculators on Tests (Teachers)	9, 13	T031701	1 Yes 2 no Omit Unmatched	11 01 01 00
Numbers and Operations Emphasis--sum of four items (Teacher)	13	T031501 T031502 T031503 T031516	Sum of codes for each: 1 Heavy 2 Moderate 3 Little 4 None Omit Unmatched	0 to 12
Teacher Questionnaire Match Status	9, 13	TCHMTCH	Full Partial Unmatched	00 10 01
Race/Ethnicity Match Status Interaction	9, 13	DRACE TCHMTCH	1 White, full 2 White, partial 3 White, unmatched 4 Black, full 5 Black, partial 6 Black, unmatched 7 Hispanic, full 8 Hispanic, partial 9 Hispanic, unmatched 10 Asian Amer., full 11 Asian Amer., partial 12 Asian Amer., unmatched 13 Other, full 14 Other, partial 15 Other, unmatched	8 0 0 0 0 0 0 0 -4 4 0 0 0 0 0 0 -4-4 0 0 0 0 0 0 -2 0 6 0 0 0 0 0 1-1-3 3 0 0 0 0 1 1-3-3 0 0 0 0 -2 0-2 0 4 0 0 0 1-1 1-1-2 2 0 0 1 1 1 1-2 2 0 0 -2 0-2 0-2 0 2 0 1-1 1-1 1-1-1 1 1 1 1 1 1 1-1-1 -2 0-2 0-2 0-2 0 1-1 1-1 1-1 1-1 1 1 1 1 1 1 1 1

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-4 (continued)

Contrast Codings for 1990 NAEP Mathematics Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Parents' Education Match Status Interaction	9, 13	PARED TCHMTCH	1 College grad, full	8 0 0 0 0 0 0 0
			2 College grad, partial	-4 4 0 0 0 0 0 0
			3 College grad, unmatched	-4-4 0 0 0 0 0 0
			4 Post-HS, full	-2 0 6 0 0 0 0 0
			5 Post-HS, partial	1-1-3 3 0 0 0 0
			6 Post-HS, unmatched	1 1-3-3 0 0 0 0
			7 HS grad, full	-2 0-2 0 4 0 0 0
			8 HS grad, partial	1-1 1-1-2 2 0 0
			9 HS grad, unmatched	1 1 1 1-2 2 0 0
			10 No HS, full	-2 0-2 0-2 0 2 0
			11 No HS, partial	1-1 1-1 1-1-1 1
			12 No HS, unmatched	1 1 1 1 1 1-1-1
			13 I don't know, full	-2 0-2 0-2 0-2 0
			14 I don't know, partial	1-1 1-1 1-1 1-1
			15 I don't know, unmatched	1 1 1 1 1 1 1 1
Region Match Status Interaction	9, 13	REGION TCHMTCH	1 Northeast, full	6 0 0 0 0 0
			2 Northeast, partial	-3 3 0 0 0 0
			3 Northeast, unmatched	-3-3 0 0 0 0
			4 Southeast, full	-2 0 4 0 0 0
			5 Southeast, partial	1-1-2 2 0 0
			6 Southeast, unmatched	1 1-2-2 0 0
			7 Central, full	-2 0-2 0 2 0
			8 Central, partial	1-1 1-1-1 1
			9 Central, unmatched	1 1 1 1-1-1
			10 West, full	-2 0-2 0-2 0
			11 West, partial	1-1 1-1 1-1
			12 West, unmatched	1 1 1 1 1 1
Gender Match Status Interaction	9, 13	DSEX TCHMTCH	1 Male, full	2 0
			2 Male, partial	-1 1
			3 Male, unmatched	-1-1
			4 Female, full	-2 0
			5 Female, partial	1-1
			6 Female, unmatched	1 1

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-5

Contrast Codings for 1990 NAEP Mathematics Trend Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Overall	All		---	1
Gender	All	DSEX	1 Male 2 Female	0 1
Observed Race/Ethnicity	All	RACE	1 White 2 Black 3 Hispanic 4 Asian American 5 American Indian 6 Other Missing	000 100 010 001 000 000 000
Size and Type of Community	All	STOC	1, 4-7 all except 2 and 3 2 Low Metro 3 High Metro	01 00 10
Region	All	REGION	1 Northeast 2 Southeast 3 Central 4 West	000 100 010 001
Parents' Education	All	PARED	1 Less than high school 2 High school graduate 3 Post-high school 4 College graduate Missing and I Don't Know	0000 1000 0100 0010 0001
Modal Grade	All	MODGRD	1 < modal grade 2 = modal grade, missing 3 > modal grade	10 00 01
Items in the Home (of newspaper, > 25 books, encyclopedia, magazines)	All	HOMEEN2	1 0 to 2 items 2 3 items 3 4 items	00 10 01

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-5 (continued)

Contrast Codings for 1990 NAEP Mathematics Trend Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Observed Race/Ethnicity by Gender ("White" includes American Indian and Other)	All	RACE DSEX	1 White, male	000
			2 Black, male	000
			3 Hispanic, male	000
			4 Asian American, male	000
			5 White, female	000
			6 Black, female	100
			7 Hispanic, female	010
			8 Asian American, female	001
Observed Race/Ethnicity by Parents' Education ("White" includes American Indian and Other)--coded differently for each age class	9	RACE PARED	1 White, < HS	0000 0000 0000
			2 White, HS graduate	0000 0000 0000
			3 White, post-HS	0000 0000 0000
			4 White, college grad.	0000 0000 0000
			5 White, missing	0000 0000 0000
			6 Black, < HS	0000 0000 0000
			7 Black, HS grad & post-HS	1000 0000 0000
			8 Black, college grad.	0010 0000 0000
			9 Black, missing	0001 0000 0000
			10 Hispanic, < HS	0000 0000 0000
			11 Hispanic, HS grad & post-HS	0000 1000 0000
			12 Hispanic, coll. grad.	0000 0010 0000
			13 Hispanic, missing	0000 0001 0000
			14 Asian Amer., < HS	0000 0000 0000
			15 Asian Amer., HS grad & post-HS	0000 0000 1000
			16 Asian Amer., coll. grad.	0000 0000 0010
			17 Asian Amer., missing	0000 0000 0001

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-5 (continued)

Contrast Codings for 1990 NAEP Mathematics Trend Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Observed Race/Ethnicity by Parents' Education ("White" includes American Indian and Other)--coded differently for each age class	13	RACE PARED	1 White, < HS	0000 0000 0000
			2 White, HS graduate	0000 0000 0000
			3 White, post-HS	0000 0000 0000
			4 White, college grad.	0000 0000 0000
			5 White, missing	0000 0000 0000
			6 Black, < HS	0000 0000 0000
			7 Black, HS graduate	1000 0000 0000
			8 Black, post-HS	0100 0000 0000
			9 Black, college grad.	0010 0000 0000
			10 Black, missing	0001 0000 0000
			11 Hispanic, < HS	0000 0000 0000
			12 Hispanic, HS grad.	0000 1000 0000
			13 Hispanic, post-HS	0000 0100 0000
			14 Hispanic, coll. grad.	0000 0010 0000
			15 Hispanic, missing	0000 0001 0000
			16 Asian Amer., < HS	0000 0000 0000
			17 Asian Amer., HS grad.	0000 0000 1000
			18 Asian Amer., post-HS	0000 0000 0100
			19 Asian Amer., coll. grad.	0000 0000 0010
			20 Asian Amer., missing	0000 0000 0001
Observed Race/Ethnicity by Parents' Education ("White" includes American Indian and Other)--coded differently for each age class	17	RACE PARED	1 White, < HS	0000 0000 0000
			2 White, HS graduate	0000 0000 0000
			3 White, post-HS	0000 0000 0000
			4 White, college grad.	0000 0000 0000
			5 White, missing	0000 0000 0000
			6 Black, < HS	0000 0000 0000
			7 Black, HS graduate	1000 0000 0000
			8 Black, post-HS	0100 0000 0000
			9 Black, college grad.	0010 0000 0000
			10 Black, missing	0001 0000 0000
			11 Hispanic, < HS	0000 0000 0000
			12 Hispanic, HS grad.	0000 1000 0000
			13 Hispanic, post-HS	0000 0100 0000
			14 Hispanic, coll. grad.	0000 0010 0000
			15 Hispanic, missing	0000 0001 0000
			16 Asian Amer., < HS	0000 0000 0000
			17 Asian Amer., HS grad.	0000 0000 1000
			18 Asian Amer., post-HS, coll. grad.	0000 0000 0100
			19 Asian Amer., missing	0000 0000 0001

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-5 (continued)

Contrast Codings for 1990 NAEP Mathematics Trend Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
School Type	All	SCHTYPE	1 Public 2 Private 3 Catholic 4 Bureau of Indian Affairs 5 Department of Defense	0 1 1 1 1
Homework	13, 17	B003901	1 None assigned 2 Didn't do 3 ½ hour or less 4 1 hour 5 2 hours 6 More than 2 hours Missing	100 010 012 013 014 000 000
Language in the Home	All	LANGHOM	1 Never 2 Sometimes 3 Always	00 10 01
Observed Race/Ethnicity by Language in the Home--coded differently for age class 9	9	RACE LANGHOM	1 White, often 2 White, sometimes 3 White, never 4 Black, often & sometimes 5 Black, never 6 Hispanic, often & sometimes 7 Hispanic, never 8 Asian Amer., often & sometimes 9 Asian Amer., never	00 00 00 00 00 00 00 00 00 10 00 00 00 00 00 00 10 00 00 00 00 00 00 10 00 00 00
Observed Race/Ethnicity by Language in the Home	13, 17	RACE LANGHOM	1 White, often 2 White, sometimes 3 White, never 4 Black, often 5 Black, sometimes 6 Black, never 7 Hispanic, often 8 Hispanic, sometimes 9 Hispanic, never 10 Asian Amer., often 11 Asian Amer., sometimes 12 Asian Amer., never	00 00 00 00 00 00 00 00 00 10 00 00 01 00 00 00 00 00 00 10 00 00 01 00 00 00 00 00 00 10 00 00 01 00 00 00

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-5 (continued)

Contrast Codings for 1990 NAEP Mathematics Trend Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Highest Level Math Taken	17	NMATH	1 Pre-algebra 2 Algebra 3 Geometry 4 Algebra 2 5 Calculus 6 Something else	10000 01000 00100 00010 00001 00000
High School Program	17	B005001	1 General 2 College preparatory 3 Vocational, technical Missing	00 10 01 00
Derived Race/Ethnicity	All	DRACE	1 White 2 Black 3 Hispanic 4 Asian American 5 Other Missing	000 100 010 001 000 000

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-6

Contrast Codings for 1990 NAEP Science Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Plant/animal Experiment	All	K810101	Yes No Missing	10 01 00
Elect. Experiment	All	K810102	Yes No Missing	10 01 00
Chem. Experiment	All	K810103	Yes No Missing	10 01 00
Rocks/min. Experiment	All	K810104	Yes No Missing	10 01 00
Telescope Experiment	All	K810105	Yes No Missing	10 01 00
Weather Experiment	All	K810106	Yes No Missing	10 01 00
Like Science	All	K810201	Yes No Missing	10 01 00
Amount of Science at School	All	K810301	Every day Several/week Once/week Less than once/week Never Missing	10000 01000 00100 00010 00001 00000
Amount of Homework Done	All	K810401	None ½ hour 1 hour 2 hours More than 2 hours No science class Missing	100000 010000 001000 000100 000010 000001 000000

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-6 (continued)

Contrast Codings for 1990 NAEP Science Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Week-long Projects	All	K810501	Yes No Missing	10 01 00
Read Textbook	All	K810601	Every day Several/week Once/week Less than once/week Never Missing	10000 01000 00100 00010 00001 00000
Discuss Science News	All	K810602	Every day Several/week Once/week Less than once/week Never Missing	10000 01000 00100 00010 00001 00000
Work with Others	All	K810603	Every day Several/week Once/week Less than once/week Never Missing	10000 01000 00100 00010 00001 00000
Give Report	All	K810604	Every day Several/week Once/week Less than once/week Never Missing	10000 01000 00100 00010 00001 00000
Do Experiments	All	K810605	Every day Several/week Once/week Less than once/week Never Missing	10000 01000 00100 00010 00001 00000

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-6 (continued)

Contrast Codings for 1990 NAEP Science Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Teacher's Lectures	13, 17	S402001	Every day	10000
			Several/week	01000
			Once/week	00100
			Less than once/week	00010
			Never	00001
			Missing	00000
Teacher Demonstrates	13, 17	S402002	Every day	10000
			Several/week	01000
			Once/week	00100
			Less than once/week	00010
			Never	00001
			Missing	00000
Teacher Asks for Reasons	13, 17	S402003	Every day	10000
			Several/week	01000
			Once/week	00100
			Less than once/week	00010
			Never	00001
			Missing	00000
Write up Experiment	13, 17	S402006	Every day	10000
			Several/week	01000
			Once/week	00100
			Less than once/week	00010
			Never	00001
			Missing	00000
Give Opinion	13, 17	S402007	Every day	10000
			Several/week	01000
			Once/week	00100
			Less than once/week	00010
			Never	00001
			Missing	00000
Use Computer	13, 17	S402008	Every day	10000
			Several/week	01000
			Once/week	00100
			Less than once/week	00010
			Never	00001
			Missing	00000

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-6 (continued)

Contrast Codings for 1990 NAEP Science Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Used a Microscope	13, 17	K810701	Yes No Missing	10 01 00
Used a Balance	13, 17	K810702	Yes No Missing	10 01 00
Used a Gas Burner	13, 17	K810703	Yes No Missing	10 01 00
Used a Wave Tank	13, 17	K810704	Yes No Missing	10 01 00
How Long Studied General Science	17	K810801	Less than 1 year 1 year ½ year Not studied Omit Not reached	10000 01000 00100 00010 00001 00000
How Long Studied Biology	17	K810802	Less than 1 year 1 year ½ year Not studied Omit Not reached	10000 01000 00100 00010 00001 00000
How Long Studied Life Science	17	K810803	Less than 1 year 1 year ½ year Not studied Omit Not reached	10000 01000 00100 00010 00001 00000
How Long Studied Chemistry	17	K810804	Less than 1 year 1 year ½ year Not studied Omit Not reached	10000 01000 00100 00010 00001 00000

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-6 (continued)

Contrast Codings for 1990 NAEP Science Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
How Long Studied Physics	17	K810805	Less than 1 year 1 year ½ year Not studied Omit Not reached	10000 01000 00100 00010 00001 00000
How Long Studied Physical Science	17	K810806	Less than 1 year 1 year ½ year Not studied Omit Not reached	10000 01000 00100 00010 00001 00000
How Long Studied Earth and Space Science	17	K810807	Less than 1 year 1 year ½ year Not studied Omit Not reached	10000 01000 00100 00010 00001 00000
Years of Teaching Science (Teacher)	13	T034701	> 20 $> 10, \leq 20$ $> 5, \leq 10$ $> 1, \leq 5$ 0, 1 Missing/not matched	10000 01000 00100 00010 00001 00000
Education Courses Taken (Teacher) (Continuous)	13	T032401	0-7 Missing/not matched	0-7 0
Education Courses Taken (Teacher) (Missing)	13	T032401	Missing Other/not matched	1 0
Biology/Life Sciences Courses Taken (Teacher) (Continuous)	13	T032402	0-7 Missing/not matched	0-7 0
Biology/Life Science Courses Taken (Teacher) (Missing)	13	T032402	Missing Other/not matched	1 0

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-6 (continued)

Contrast Codings for 1990 NAEP Science Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Chemistry Courses Taken (Teacher) (Continuous)	13	T032403	0-7 Missing/not matched	0-7 0
Chemistry Courses Taken (Teacher) (Missing)	13	T032403	Missing Other/not matched	1 0
Physics Courses Taken (Teacher) (Continuous)	13	T032404	0-7 Missing/not matched	0-7 0
Physics Courses Taken (Teacher) (Missing)	13	T032404	Missing Other/not matched	1 0
Earth Sciences Courses Taken (Teacher) (Continuous)	13	T032405	0-7 Missing/not matched	0-7 0
Earth Sciences Courses Taken (Teacher) (Missing)	13	T032405	Missing Other/not matched	1 0
Adequate Laboratory Facilities (Teacher)	13	T032902	Strongly agree Agree No opinion Disagree Strongly Disagree Missing/not matched	100 100 010 001 001 000
Well-supplied Instructional Materials (Teacher)	13	T032903	Strongly agree Agree No opinion Disagree Strongly Disagree Missing/not matched	100 100 010 001 001 000
Textbook Curriculum (Teacher)	13	T032905	Strongly agree Agree No opinion Disagree Strongly Disagree Missing/not matched	100 100 010 001 001 000

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-6 (continued)

Contrast Codings for 1990 NAEP Science Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Emphasis on Knowing Facts (Teacher)	13	T033401	Heavy Moderate Little None Missing/not matched	1000 0100 0010 0001 0000
Emphasis on Key Concepts (Teacher)	13	T033402	Heavy Moderate Little None Missing/not matched	1000 0100 0010 0001 0000
Emphasis on Problem Solving (Teacher)	13	T033403	Heavy Moderate Little None Missing/not matched	1000 0100 0010 0001 0000
Emphasis on Nature of Science (Teacher)	13	T033405	Heavy Moderate Little None Missing/not matched	1000 0100 0010 0001 0000
Emphasis on Lab Techniques (Teacher)	13	T033408	Heavy Moderate Little None Missing/not matched	1000 0100 0010 0001 0000
Ability of Class (Teacher)	13	T033701	High Average Low Mixed Missing/not matched	100 010 001 010 000
Content of Course (Teacher)	13	T033801	General science Life science Earth science Physical science Integrated science Missing/not matched	10000 01000 00100 00010 00001 00000

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-6 (continued)

Contrast Codings for 1990 NAEP Science Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Lecture (Teacher)	13	T034101	Almost every day Several times/week Once/week Less than once/week Never Missing/not matched	1000 0100 0010 0010 0001 0000
Read Textbook (Teacher)	13	T034103	Almost every day Several times/week Once/week Less than once/week Never Missing/not matched	1000 0100 0010 0010 0001 0000
Use Computer (Teacher)	13	T034110	Almost every day Several times/week Once/week Less than once/week Never Missing/not matched	1000 0100 0010 0010 0001 0000
Level of Science Activities (Teacher): Total of scores (0, 1, 2) for six items	13	T034102 T034105 T034106 T034107 T034108 T034109	Continuous Almost every day Several times/week Once/week Less than once/week Never Missing/not matched	0-12 (2) (2) (1) (1) (0) (0)
Time Spent Weekly on Homework (Teacher)	13	T034301	None ½ hour 1 hour 2 hours More than 2 hours Missing/not matched	0 1 2 3 4 0
Teacher Questionnaire Match Status	13	TCHMTCH	Full Partial Unmatched	00 10 01

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-6 (continued)

Contrast Codings for 1990 NAEP Science Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Race/Ethnicity Match Status Interaction	13	DRACE TCHMTCH	1 White, full	8 0 0 0 0 0 0 0
			2 White, partial	-4 4 0 0 0 0 0 0
			3 White, unmatched	-4-4 0 0 0 0 0 0
			4 Black, full	-2 0 6 0 0 0 0 0
			5 Black, partial	1-1-3 3 0 0 0 0
			6 Black, unmatched	1 1-3-3 0 0 0 0
			7 Hispanic, full	-2 0-2 0 4 0 0 0
			8 Hispanic, partial	1-1 1-1-2 2 0 0
			9 Hispanic, unmatched	1 1 1 1-2 2 0 0
			10 Asian Amer., full	-2 0-2 0-2 0 2 0
			11 Asian Amer., partial	1-1 1-1 1-1-1 1
			12 Asian Amer., unmatched	1 1 1 1 1 1-1-1
			13 Other, full	-2 0-2 0-2 0-2 0
			14 Other, partial	1-1 1-1 1-1 1-1
			15 Other, unmatched	1 1 1 1 1 1 1 1
Parents' Education Match Status Interaction	13	PARED TCHMTCH	1 College grad, full	8 0 0 0 0 0 0 0
			2 College grad, partial	-4 4 0 0 0 0 0 0
			3 College grad, unmatched	-4-4 0 0 0 0 0 0
			4 Post-HS, full	-2 0 6 0 0 0 0 0
			5 Post-HS, partial	1-1-3 3 0 0 0 0
			6 Post-HS, unmatched	1 1-3-3 0 0 0 0
			7 HS grad, full	-2 0-2 0 4 0 0 0
			8 HS grad, partial	1-1 1-1-2 2 0 0
			9 HS grad, unmatched	1 1 1 1-2 2 0 0
			10 No HS, full	-2 0-2 0-2 0 2 0
			11 No HS, partial	1-1 1-1 1-1-1 1
			12 No HS, unmatched	1 1 1 1 1 1-1-1
			13 I don't know, full	-2 0-2 0-2 0-2 0
			14 I don't know, partial	1-1 1-1 1-1 1-1
			15 I don't know, unmatched	1 1 1 1 1 1 1 1

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-6 (continued)

Contrast Codings for 1990 NAEP Science Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Region Match Status Interaction	13	REGION TCHMTCH	1 Northeast, full	6 0 0 0 0 0
			2 Northeast, partial	-3 3 0 0 0 0
			3 Northeast, unmatched	-3-3 0 0 0 0
			4 Southeast, full	-2 0 4 0 0 0
			5 Southeast, partial	1-1-2 2 0 0
			6 Southeast, unmatched	1 1-2-2 0 0
			7 Central, full	-2 0-2 0 2 0
			8 Central, partial	1-1 1-1-1 1
			9 Central, unmatched	1 1 1 1-1-1
			10 West, full	-2 0-2 0-2 0
			11 West, partial	1-1 1-1 1-1
			12 West, unmatched	1 1 1 1 1 1
Gender Match Status Interaction	13	DSEX TCHMTCH	1 Male, full	2 0
			2 Male, partial	-1 1
			3 Male, unmatched	-1-1
			4 Female, full	-2 0
			5 Female, partial	1-1
			6 Female, unmatched	1 1

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-7

Contrast Codings for 1990 NAEP Science Trend Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Overall	All		---	1
Gender	All	DSEX	1 Male 2 Female	0 1
Observed Race/Ethnicity	All	RACE	1 White 2 Black 3 Hispanic 4 Asian American 5 American Indian 6 Other Missing	000 100 010 001 000 000 000
Size and Type of Community	All	STOC	2 Low Metro 3 High Metro 1, 4-7 All others and Missing	10 01 00
Region	All	REGION	1 Northeast 2 Southeast 3 Central 4 West	000 100 010 001
Parents' Education	All	PARED	1 Less than high school 2 High school graduate 3 Post-high school 4 College graduate Missing and I Don't Know	0000 1000 0100 0010 0001
Modal Grade	All	MODGRD	1 < modal grade 2 = modal grade, missing 3 > modal grade	10 00 01
Items in the Home (of newspaper, > 25 books, encyclopedia, magazines)	All	HOMEEN2	1 0 to 2 items 2 3 items 3 4 items	00 10 01

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-7 (continued)

Contrast Codings for 1990 NAEP Science Trend Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Observed Race/Ethnicity by Gender ("White" includes American Indian and Other)	All	RACE DSEX	1 White, male	000
			2 Black, male	000
			3 Hispanic, male	000
			4 Asian American, male	000
			5 White, female	000
			6 Black, female	100
			7 Hispanic, female	010
			8 Asian American, female	001
			Missing	000
Observed Race/Ethnicity by Parents' Education ("White" includes American Indian and Other)	All	RACE PARED	1 White, < HS	0000 0000 0000
			2 White, HS graduate	0000 0000 0000
			3 White, post-HS	0000 0000 0000
			4 White, college grad.	0000 0000 0000
			5 White, missing	0000 0000 0000
			6 Black, < HS	0000 0000 0000
			7 Black, HS graduate	1000 0000 0000
			8 Black, post-HS	0100 0000 0000
			9 Black, college grad.	0010 0000 0000
			10 Black, missing	0001 0000 0000
			11 Hispanic, < HS	0000 0000 0000
			12 Hispanic, HS grad.	0000 1000 0000
			13 Hispanic, post-HS	0000 0100 0000
			14 Hispanic, coll. grad.	0000 0010 0000
			15 Hispanic, missing	0000 0001 0000
			16 Asian Amer., < HS	0000 0000 0000
			17 Asian Amer., HS grad.	0000 0000 1000
			18 Asian Amer., post-HS	0000 0000 0100
			19 Asian Amer., coll. grad.	0000 0000 0010
			20 Asian Amer., missing	0000 0000 0001
School Type	All	SCHTYPE	1 Public	0
			2 Private	1
			3 Catholic	1
			4 Bureau of Indian Affairs	1
			5 Department of Defense	1

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-7 (continued)

Contrast Codings for 1990 NAEP Science Trend Conditioning Variables

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Homework	13, 17	B003901	1 None assigned 2 Didn't do 3 ½ hour or less 4 1 hour 5 2 hours 6 More than 2 hours Missing	100 010 012 013 014 000 000
Language in the Home	All	LANGHOM	1 Never 2 Sometimes 3 Always	00 10 01
Observed Race/Ethnicity by Language in the Home	All	RACE LANGHOM	1 White, often 2 White, sometimes 3 White, never 4 Black, often 5 Black, sometimes 6 Black, never 7 Hispanic, often 8 Hispanic, sometimes 9 Hispanic, never 10 Asian Amer., often 11 Asian Amer., sometimes 12 Asian Amer., never	00 00 00 00 00 00 00 00 00 10 00 00 01 00 00 00 00 00 00 10 00 00 01 00 00 00 00 00 00 10 00 00 01 00 00 00
Number of Science Courses	17	NSCI	1 General science 2 Biology 3 Chemistry 4 Physics 5 Nothing/something else	1000 0100 0010 0001 0000
High School Program	17	B005001	1 General 2 College preparatory 3 Vocational, technical Missing	00 10 01 00
Derived Race/Ethnicity	All	DRACE	1 White 2 Black 3 Hispanic 4 Asian American 5 Other Missing	000 100 010 001 000 000

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table F-8
Estimated Effects for Reading Cross-sectional Conditioning Variable Contrasts, Age 9/Grade 4

Contrast	Estimated Effect	Description
1 OVERALL	-1.173283	OVERALL CONSTANT
2 GENDER2	0.135073	SEX (FEMALE)
3 ETHNIC2	-0.298142	DERIVED RACE (BLACK)
4 ETHNIC3	-0.268280	DERIVED RACE (HISPANIC)
5 ETHNIC4	-0.053435	DERIVED RACE (ASIAN)
6 STOC3	0.023199	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7 STOC147	-0.019137	SIZE AND TYPE OF COMMUNITY (NON-HI&LOW)
8 REGION2	0.008789	REGION (SOUTHEAST)
9 REGION3	-0.001665	REGION (CENTRAL)
10 REGION4	0.055805	REGION (WEST)
11 PARED2	0.086396	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12 PARED3	0.263410	PARENTS EDUCATION (POST HIGH SCHOOL)
13 PARED4	0.233996	PARENTS EDUCATION (COLLEGE GRAD)
14 PARED_	0.114985	PARENTS EDUCATION (MISSING,I DON'T KNOW)
15 HOMEITM2	0.103797	3 ITEMS IN THE HOME
16 HOMEITM3	0.166830	4 ITEMS IN THE HOME
17 TVWATCHL	0.177517	# HOURS OF DAILY TV VIEWING (LINEAR)
18 TVWATCHQ	-0.030460	# HOURS OF DAILY TV VIEWING (QUADRATIC)
19 LANGHM23	-0.053267	LANG OTHER THAN ENG AT HOME(SOME/ALWAYS)
20 HW-NO	-0.104386	HOMEWORK (NONE ASSIGNED)
21 HW-YES	-0.442067	HOMEWORK (YES, ASSIGNED)
22 HWLINEAR	0.495070	HOMEWORK (LINEAR)
23 HW:QUAD	-0.136699	HOMEWORK (QUADRATIC)
24 PCTLUNCH	-0.000227	PERCENT IN LUNCH PROGRAM (LINEAR)
25 %LUNCH M	0.010882	PERCENT IN LUNCH PROGRAM - MISSING
26 PCTWHT1	0.175985	PERCENT WHITE IN SCHOOL (0-49 %) WHT MIN
27 PCTWHT2	0.066838	PERCENT WHITE IN SCHOOL (50-79%) INTGRAT
28 AGE/GRD2	-0.422535	AGE X GRADE: MODAL AGE, < MODAL GRADE
29 AGE/GRD3	0.088440	AGE X GRADE: MODAL AGE, MODAL GRADE,MISS
30 AGE/GRD4	0.355574	AGE X GRADE: MODAL AGE, > MODAL GRADE
31 AGE/GRD5	-0.019680	AGE X GRADE: > MODAL AGE, MODAL GRADE
32 SCHTYPE	-0.016771	SCHOOL TYPE: NON-PUBLIC
33 HW HELP	-0.127309	SOMEONE AT HOME HELPS WITH HW:>ONCE WEEK
34 PRESCH_Y	0.060268	WENT TO PRESCHOOL? (YES)
35 SINGLEP1	0.055476	HOW MANY PARENTS AT HOME? (BOTH)
36 MOMHOME1	0.345018	DOES YOUR MOTHER LIVE AT HOME? (YES)
37 MOMWORKY	-0.045070	DOES YOUR MOTHER WORK FOR PAY? (YES)
38 PGS RD14	0.055266	PAGES/DAY READ FOR SCHOOL (6 OR MORE)
39 PGS RD13	-0.009898	PAGES/DAY READ FOR SCHOOL (11 OR MORE)
40 SCHMATH	-0.029615	SCHOOL LEVEL PROFICIENCY - MATH
41 SCHMATHM	0.006182	SCHOOL LEVEL PROFICIENCY MATH (MISSING)
42 SCHREAD	0.289057	SCHOOL LEVEL PROFICIENCY - READING
43 SCHREADM	-0.799751	SCHOOL LEVEL PROFICIENCY READING (MISSG)
44 SCHSCI	0.035248	SCHOOL LEVEL PROFICIENCY - SCIENCE
45 SCHSCIM	0.672003	SCHOOL LEVEL PROFICIENCY SCIENCE (MISSG)
46 SEASON_W	-0.027801	SEASON - WINTER
47 READOUTS	-0.029322	DO YOU READ OUTSIDE OF SCHOOL (LIN)
48 TELLABT1	0.173826	DO YOU TELL OTHERS ABOUT BOOK (LIN)
49 TELLABT2	-0.032749	DO YOU TELL OTHERS ABOUT BOOK (QUA)

Table F-8 (continued)
Estimated Effects for Reading Cross-sectional Conditioning Variable Contrasts, Age 9/Grade 4

Contrast	Estimated Effect	Description
50 LIBRARYL	0.316772	DO YOU TAKE A LIB. BOOK FOR SELF (LIN)
51 LIBRARYQ	-0.056465	DO YOU TAKE A LIB. BOOK FOR SELF (QUA)
52 TECHWORD	0.002555	DOES TEACHER TEACH NEW WORDS (LIN)
53 TECHWORD	0.002905	DOES TEAHCER TEACH NEW WORDS (QUA)
54 ASKABOUT	-0.048940	TEACHER ASKS ABOUT BOOK (LIN)
55 ASKABOUT	0.007929	TEACHER ASKS ABOUT BOOK (QUA)
56 WORKBOKL	-0.023038	DO YOU WORK IN WORKBOOK (LIN)
57 WORKBOKQ	-0.001267	DO YOU WORK IN WORKBOOK (QUA)
58 WRITEABL	0.160037	DO YOU WRITE ABOUT BOOK (LIN)
59 WRITEABQ	-0.022990	DO YOU WRITE ABOUT BOOK (QUA)
60 GROUPL	0.242211	DO YOU DO A GROUP ACTIVITY (LIN)
61 GROUPQ	-0.026912	DO YOU DO A GROUP ACTIVITY (QUA)
62 READER	-0.153706	WHAT KIND OF READER ARE YOU (LIN)

Table F-9

Estimated Effects for Reading Cross-sectional Conditioning Variable Contrasts, Age 13/Grade 8

Contrast	Estimated Effect	Description
1 OVERALL	-1.347542	OVERALL CONSTANT
2 GENDER2	0.168486	SEX (FEMALE)
3 ETHNIC2	-0.243906	DERIVED RACE (BLACK)
4 ETHNIC3	-0.236634	DERIVED RACE (HISPANIC)
5 ETHNIC4	0.033875	DERIVED RACE (ASIAN)
6 STOC3	-0.003784	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7 STOC147	0.034684	SIZE AND TYPE OF COMMUNITY (NON-HI&LOW)
8 REGION2	0.043577	REGION (SOUTHEAST)
9 REGION3	-0.001407	REGION (CENTRAL)
10 REGION4	0.084632	REGION (WEST)
11 PARED2	0.021343	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12 PARED3	0.155018	PARENTS EDUCATION (POST HIGH SCHOOL)
13 PARED4	0.222364	PARENTS EDUCATION (COLLEGE GRAD)
14 PARED_	0.012097	PARENTS EDUCATION (MISSING,I DON'T KNOW)
15 HOMEITM2	0.051685	3 ITEMS IN THE HOME
16 HOMEITM3	0.122863	4 ITEMS IN THE HOME
17 TVWATCHL	0.014105	# HOURS OF DAILY TV VIEWING (LINEAR)
18 TVWATCHQ	-0.004682	# HOURS OF DAILY TV VIEWING (QUADRATIC)
19 LANGHM23	-0.005874	LANG OTHER THAN ENG AT HOME(SOME/ALWAYS)
20 HW-NO	0.[illegible]001	HOMEWORK (NONE ASSIGNED)
21 HW-YES	0.0[illegible]8270	HOMEWORK (YES, ASSIGNED)
22 HWLINEAR	0.128345	HOMEWORK (LINEAR)
23 HW:QUAD	-0.024579	HOMEWORK (QUADRATIC)
24 PCTLUNCH	0.000762	PERCENT IN LUNCH PROGRAM (LINEAR)
25 %LUNCH M	0.008189	PERCENT IN LUNCH PROGRAM - MISSING
26 PCTWHT[illegible]	0.113835	PERCENT WHITE IN SCHOOL (0-49 %) WHT MIN
27 PCTWHT[illegible]	0.058156	PERCENT WHITE IN SCHOOL (50-79%) INTGRAT
28 AGE/GRD2	-0.497860	AGE X GRADE: MODAL AGE, < MODAL GRADE
29 AGE/GRD3	-0.163393	AGE X GRADE: MODAL AGE, MODAL GRADE,MISS
30 AGE/GRD4	0.582340	AGE X GRADE: MODAL AGE, > MODAL GRADE
31 AGE/GRD5	-0.333893	AGE X GRADE: > MODAL AGE, MODAL GRADE
32 SCHTYPE	-0.049750	SCHOOL TYPE: NON-PUBLIC
33 HW HELP	-0.194827	SOMEONE AT HOME HELPS WITH HW:>ONCE WEEK
34 SINGLEP1	0.049648	HOW MANY PARENTS AT HOME? (BOTH)
35 MOMHOME1	0.063634	DOES YOUR MOTHER LIVE AT HOME? (YES)
36 MOMWORKY	-0.069059	DOES YOUR MOTHER WORK FOR PAY? (YES)
37 PGS RD14	0.105463	PAGES/DAY READ FOR SCHOOL (6 OR MORE)
38 PGS RD13	0.023371	PAGES/DAY READ FOR SCHOOL (11 OR MORE)
39 EXPHSGRD	0.210213	EXPECT TO GRADUATE FROM HIGH SCHOOL(YES)
40 SCHMSS12	0.122226	SCHOOL DAYS MISSED LAST MONTH (0-2 DAYS)
41 BEHAVRNM	-0.218580	RULES OF BEHAVIOR ARE STRICT (NON-MISSG)
42 BEHAVR-L	0.032516	RULES OF BEHAVIOR ARE STRICT (LINEAR)
43 SAFETYNM	-0.189244	DON'T FEEL SAFE AT SCHOOL (NON-MISSG)
44 SAFETY-L	0.102113	DON'T FEEL SAFE AT SCHOOL (LINEAR)
45 DISRPTNM	-0.231707	STUDENTS OFTEN DISRUPT CLASS (NON-MISSG)
46 DISRPT-L	0.011902	STUDENTS OFTEN DISRUPT CLASS (LINEAR)
47 SCHMATH	-0.024198	SCHOOL LEVEL PROFICIENCY - MATH
48 SCHMATHM	-0.003961	SCHOOL LEVEL PROFICIENCY MATH (MISSING)
49 SCHREAD	0.312821	SCHOOL LEVEL PROFICIENCY - READING

Table F-9 (continued)
Estimated Effects for Reading Cross-sectional Conditioning Variable Contrasts, Age 13/Grade 8

	Contrast	Estimated Effect	Description
50	SCHREADM	0.103545	SCHOOL LEVEL PROFICIENCY READING (MISSG)
51	SCHSCI	-0.014015	SCHOOL LEVEL PROFICIENCY - SCIENCE
52	SCHSCIM	-0.293727	SCHOOL LEVEL PROFICIENCY SCIENCE (MISSG)
53	SEASON_W	-0.011387	SEASON - WINTER=1 SPRING=2 (LIN)
54	OFTNREAD	-0.071278	HOW OFTEN READ WHAT YOU WANT (LIN)
55	TELLPEPL	0.209591	HOW OFTN TELL PEPLE WHAT YOU READ (LIN)
56	TELLPEPQ	-0.036543	HOW OFTN TELL PEPLE WHAT YOU READ (QUA)
57	LIBRARYL	0.209898	HOW OFTEN TAKE BOOKS FROM LIBRARY (LIN)
58	LIBRARYQ	-0.029016	HOW OFTEN TAKE BOOKS FROM LIBRARY (QUA)
59	AUTHOR+	-0.034951	HOW OFTEN READ MORE BY AUTHOR (LIN)
60	NEWVOCAL	0.020954	HOW OFTEN DISCUSS NEW VOCABULARY (LIN)
61	NEWVOCAQ	-0.011524	HOW OFTEN DISCUSS NEW VOCABULARY (QUA)
62	TEACHASK	-0.050214	HOW OFTEN TEACHER ASK ABOUT BOOK (LIN)
63	WORKBOKL	0.067089	HOW OFTEN WORK IN WORKBOOK (LIN)
64	WORKBOKQ	-0.014679	HOW OFTEN WORK IN WORKBOOK (QUA)
65	JOURNAL	0.048040	HOW OFTEN WRITE IN JOURNAL (LIN)
66	REPORTL	0.289088	HOW OFTN WRITE A REPORT ABOUT BOOK (LIN)
67	REPORTQ	-0.041664	HOW OFTN WRITE A REPORT ABOUT BOOK (QUA)
68	ACTIVITL	0.279719	HOW OFTN DO AN ACTIVITY ABOUT BOOK (LIN)
69	ACTIVITQ	-0.033727	HOW OFTN DO AN ACTIVITY ABOUT BOOK (QUA)
70	THINKL	-0.067260	HOW OFTN DO YOU THINK ABOUT AUTHOR (LIN)
71	THINKQ	0.018356	HOW OFTN DO YOU THINK ABOUT AUTHOR (QUA)
72	EXPLAIN	-0.025123	HOW OFTN EXPLN YOUR UNDERSTANDING (LIN)
73	DISCUSS	0.009022	HOW OFTEN DISC DIFF INTERPRETATION(LIN)
74	PREDICTL	0.043231	HOW OFTEN PREDICT WHAT YOU FIND (LIN)
75	PREDICTQ	-0.006266	HOW OFTEN PREDICT WHAT YOU FIND (QUA)
76	READSPAR	0.025943	WHAT KIND OF READING IN SPARE TIME (LIN)
77	ADLTREAD	0.034837	WHAT KIND OF READING DO ADULTS DO (LIN)
78	KINDOFRE	-0.172974	WHAT KIND OF READER DO YOU THINK YOU ARE

Table F-10
Estimated Effects for Reading Cross-sectional Conditioning Variable Contrasts, Age 17/Grade 12

Contrast	Estimated Effect	Description
1 OVERALL	-1.176652	OVERALL CONSTANT
2 GENDER2	0.073004	SEX (FEMALE)
3 ETHNIC2	-0.310819	DERIVED RACE (BLACK)
4 ETHNIC3	-0.182205	DERIVED RACE (HISPANIC)
5 ETHNIC4	-0.102808	DERIVED RACE (ASIAN)
6 STOC3	-0.127369	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7 STOC147	-0.107488	SIZE AND TYPE OF COMMUNITY (NON-HI&LOW)
8 REGION2	0.073154	REGION (SOUTHEAST)
9 REGION3	0.033297	REGION (CENTRAL)
10 REGION4	0.060369	REGION (WEST)
11 PARED2	-0.034460	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12 PARED3	0.080373	PARENTS EDUCATION (POST HIGH SCHOOL)
13 PARED4	0.059344	PARENTS EDUCATION (COLLEGE GRAD)
14 PARED_	-0.166371	PARENTS EDUCATION (MISSING,I DON'T KNOW)
15 HOMEITM2	0.088221	3 ITEMS IN THE HOME
16 HOMEITM3	0.067119	4 ITEMS IN THE HOME
17 TVWATCHL	-0.016828	# HOURS OF DAILY TV VIEWING (LINEAR)
18 TVWATCHQ	0.001519	# HOURS OF DAILY TV VIEWING (QUADRATIC)
19 LANGHM23	-0.018812	LANG OTHER THAN ENG AT HOME(SOME/ALWAYS)
20 HW-NO	0.102385	HOMEWORK (NONE ASSIGNED)
21 HW-YES	0.104887	HOMEWORK (YES, ASSIGNED)
22 HWLINEAR	0.058159	HOMEWORK (LINEAR)
23 HW:QUAD	-0.008312	HOMEWORK (QUADRATIC)
24 PCTLUNCH	0.001654	PERCENT IN LUNCH PROGRAM (LINEAR)
25 %LUNCH M	-0.006105	PERCENT IN LUNCH PROGRAM - MISSING
26 PCTWHT1	0.095737	PERCENT WHITE IN SCHOOL (0-49 %) WHT MIN
27 PCTWHT2	0.032876	PERCENT WHITE IN SCHOOL (50-79%) INTGRAT
28 AGE/GRD2	-0.228654	AGE X GRADE: MODAL AGE, < MODAL GRADE
29 AGE/GRD3	-0.022832	AGE X GRADE: MODAL AGE, MODAL GRADE,MISS
30 AGE/GRD4		AGE X GRADE: MODAL AGE, > MODAL GRADE
31 AGE/GRD5	-0.144393	AGE X GRADE: > MODAL AGE, MODAL GRADE
32 SCHTYPE	-0.153433	SCHOOL TYPE: NON-PUBLIC
33 HW HELP	-0.238052	SOMEONE AT HOME HELPS WITH HW:>ONCE WEEK
34 SINGLEP1	0.011151	HOW MANY PARENTS AT HOME? (BOTH)
35 MOMHOME1	0.042361	DOES YOUR MOTHER LIVE AT HOME? (YES)
36 MOMWORKY	-0.018263	DOES YOUR MOTHER WORK FOR PAY? (YES)
37 PGS RD14	0.060373	PAGES/DAY READ FOR SCHOOL (6 OR MORE)
38 PGS RD13	0.034984	PAGES/DAY READ FOR SCHOOL (11 OR MORE)
39 SCHMSS12	0.078387	SCHOOL DAYS MISSED LAST MONTH (0-2 DAYS)
40 HS PROG2	0.171681	HIGH SCHOOL PROGRAM (2-COLLEGE PREP)
41 HS PROG3	-0.133079	HIGH SCHOOL PROGRAM (3-VOCATNL,TECHNCL)
42 POST HS2	0.007352	POST-SECONDARY PLANS (2-TWO-YR COLLEGE)
43 POST HS3	0.171938	POST-SECONDARY PLANS (3-FOUR-YR COLLEGE)
44 BEHAVRNM	-0.171179	RULES OF BEHAVIOR ARE STRICT (NON-MISSG)
45 BEHAVR-L	0.044522	RULES OF BEHAVIOR ARE STRICT (LINEAR)
46 SAFETYNM	-0.158600	DON'T FEEL SAFE AT SCHOOL (NON-MISSG)
47 SAFETY-L	0.090651	DON'T FEEL SAFE AT SCHOOL (LINEAR)
48 DISRPTNM	-0.088038	STUDENTS OFTEN DISRUPT CLASS (NON-MISSG)
49 DISRPT-L	0.025039	STUDENTS OFTEN DISRUPT CLASS (LINEAR)

Table F-10 (continued)
Estimated Effects for Reading Cross-sectional Conditioning Variable Contrasts, Age 17/Grade 12

Contrast	Estimated Effect	Description
50 SEMENGNM	-0.181925	SEMESTERS OF ENGLSH/LIT/WRITE (NON-MISS)
51 SEMENG-L	0.049688	SEMESTERS OF ENGLSH/LIT/WRITE (LINEAR)
52 SEMMTHNM	-0.038389	SEMESTERS OF MATHEMATICS (NON-MISSING)
53 SEMMTH-L	0.006465	SEMESTERS OF MATHEMATICS (LINEAR)
54 SEMSCINM	-0.153716	SEMESTERS OF SCIENCE (NON-MISSING)
55 SEMSCI-L	0.031296	SEMESTERS OF SCIENCE (LINEAR)
56 SEMHISNM	-0.049017	SEMESTERS OF HIST/SS/GEOGRPHY (NON-MISS)
57 SEMHIS-L	0.007155	SEMESTERS OF HIST/SS/GEOGRPHY (LINEAR)
58 SEMFLGNM	-0.025497	SEMESTERS OF FOREIGN LANGUAGE (NON-MISS)
59 SEMFLG-L	0.003015	SEMESTERS OF FOREIGN LANGUAGE (LINEAR)
60 SEMVOCNM	0.076056	SEMESTERS OF VOC/TCH/BUS ED (NON-MISS)
61 SEMVOC-L	-0.020378	SEMESTERS OF VOC/TCH/BUS ED (LINEAR)
62 SEMARTNM	-0.068883	SEMESTERS OF ART/MUSIC (NON-MISSING)
63 SEMART-L	-0.005105	SEMESTERS OF ART/MUSIC (LINEAR)
64 SCHMATH	-0.061758	SCHOOL LEVEL PROFICIENCY - MATH
65 SCHMATHM	0.057551	SCHOOL LEVEL PROFICIENCY MATH (MISSING)
66 SCHREAD	0.247908	SCHOOL LEVEL PROFICIENCY - READING
67 SCHREADM	-1.080290	SCHOOL LEVEL PROFICIENCY READING (MISSG)
68 SCHSCI	0.062414	SCHOOL LEVEL PROFICIENCY - SCIENCE
69 SCHSCIM	0.224779	SCHOOL LEVEL PROFICIENCY SCIENCE (MISSG)
70 SEASON_W	0.026092	SEASON - WINTER=1 SPRING=2 (LIN)
71 OFTNREAD	-0.021980	HOW OFTEN READ WHAT YOU WANT (LIN)
72 TELLPEPL	0.085060	HOW OFTN TELL PEOPLE WHAT YOU READ (LIN)
73 TELLPEPQ	-0.020693	HOW OFTN TELL PEOPLE WHAT YOU READ (QUA)
74 LIBRARYL	0.245155	HOW OFTEN TAKE BOOKS FROM LIBRARY (LIN)
75 LIBRARYQ	-0.035969	HOW OFTEN TAKE BOOKS FROM LIBRARY (QUA)
76 AUTHOR+	-0.048550	HOW OFTEN READ MORE BY AUTHOR (LIN)
77 NEWVOCAL	0.020365	HOW OFTEN DISCUSS NEW VOCABULARY (LIN)
78 NEWVOCAQ	-0.007839	HOW OFTEN DISCUSS NEW VOCABULARY (QUA)
79 TEACHASK	-0.061795	HOW OFTEN TEACHER ASK ABOUT BOOK (LIN)
80 WORKBOKL	0.048252	HOW OFTEN WORK IN WORKBOOK (LIN)
81 WORKBOKQ	-0.007205	HOW OFTEN WORK IN WORKBOOK (QUA)
82 JOURNAL	0.063823	HOW OFTEN WRITE IN JOURNAL (LIN)
83 REPORTL	0.230938	HOW OFTN WRITE A REPORT ABOUT BOOK (LIN)
84 REPORTQ	-0.035896	HOW OFTN WRITE A REPORT ABOUT BOOK (QUA)
85 ACTIVITL	0.118669	HOW OFTN DO AN ACTIVITY ABOUT BOOK (LIN)
86 ACTIVITQ	-0.013688	HOW OFTN DO AN ACTIVITY ABOUT BOOK (QUA)
87 THINKL	-0.046179	HOW OFTN DO YOU THINK ABOUT AUTHOR (LIN)
88 THINKQ	0.012173	HOW OFTN DO YOU THINK ABOUT AUTHOR (QUA)
89 EXPLAIN	-0.060474	HOW OFTN EXPLAIN YOUR UNDSTNDING (LIN)
90 DISCUSS	-0.008209	HOW OFTN DISCUSS DIFF INTP. (LIN)
91 PREDICTL	-0.064962	HOW OFTEN PREDICT WHAT YOU FIND (LIN)
92 PREDICTQ	0.013137	HOW OFTEN PREDICT WHAT YOU FIND (QUA)
93 NEWINFOL	0.025432	TEACHER ASK TO READ FOR NEW INFO (LIN)
94 NEWINFOQ	-0.003485	TEACHER ASK TO READ FOR NEW INFO (QUA)
95 EXTENDL	-0.026881	TEACHER ASK TO EXTEND LIT. IDEAS (LIN)
96 EXTENDQ	0.004356	TEACHER ASK TO EXTEND LIT. IDEAS (QUA)
97 SKILLSL	0.015337	TEACHER SHOW HOW TO USE READ SKILLS (LIN
98 SKILLSQ	0.002488	TEACHER SHOW TO USE READ SKILLS (QUA)

Table F-10 (continued)
Estimated Effects for Reading Cross-sectional Conditioning Variable Contrasts, Age 17/Grade 12

	Contrast	Estimated Effect	Description
99	TEXTBOKL	-0.058974	HOW OFTEN READ TEXTBOOK FOR ASSIGN (LIN)
100	NEWSMAGL	0.115967	HOW OFTEN READ NEWS/MAG FOR ASSIGN (LIN)
101	NEWSMAGQ	-0.017200	HOW OFTEN READ NEWS/MAG FOR ASSIGN (QUA)
102	STORYL	-0.017516	HOW OFTN READ STORY/POEM 4 ASSIGN (LIN)
103	ENCYLPDL	0.033018	HOW OFTN READ ENCYL/DICT 4 ASSIGN (LIN)
104	ENCYLPDQ	-0.000285	HOW OFTN READ ENCYL/DICT 4 ASSIGN (QUA)
105	READSPAR	0.029378	WHAT KIND OF READING IN SPARE TIME (LIN)
106	ADLTREAD	0.048895	WHAT KIND OF READING DO ADULTS DO (LIN)
107	KINDOFRE	-0.117349	WHAT KIND OF READER DO YOU THINK YOU ARE

Table F-11

Estimated Effects for Reading Answer Booklet Bridge Conditioning Variable Contrasts, Age 9/Grade 4

Contrast	Estimated Effect	Description
1 OVERALL	-2.056382	OVERALL CONSTANT
2 GENDER2	0.079995	SEX (FEMALE)
3 ETHNIC2	-0.138825	DERIVED RACE (BLACK)
4 ETHNIC3	-0.229808	DERIVED RACE (HISPANIC)
5 ETHNIC4	-0.023641	DERIVED RACE (ASIAN)
6 STOC3	-0.183828	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7 STOC147	-0.047934	SIZE AND TYPE OF COMMUNITY (NON-HI&LOW)
8 REGION2	-0.042043	REGION (SOUTHEAST)
9 REGION3	0.011633	REGION (CENTRAL)
10 REGION4	-0.024295	REGION (WEST)
11 PARED2	-0.001549	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12 PARED3	0.239846	PARENTS EDUCATION (POST HIGH SCHOOL)
13 PARED4	0.145079	PARENTS EDUCATION (COLLEGE GRAD)
14 PARED_	0.012283	PARENTS EDUCATION (MISSING,I DON'T KNOW)
15 HOMEITM2	0.076833	3 ITEMS IN THE HOME
16 HOMEITM3	0.057936	4 ITEMS IN THE HOME
17 TVWATCHL	0.092674	# HOURS OF DAILY TV VIEWING (LINEAR)
18 TVWATCHQ	-0.019943	# HOURS OF DAILY TV VIEWING (QUADRATIC)
19 LANGHM23	-0.027031	LANG OTHER THAN ENG AT HOME(SOME/ALWAYS)
20 HW-NO	-0.062030	HOMEWORK (NONE ASSIGNED)
21 HW-YES	-0.311001	HOMEWORK (YES, ASSIGNED)
22 HWLINEAR	0.414979	HOMEWORK (LINEAR)
23 HW:QUAD	-0.121715	HOMEWORK (QUADRATIC)
24 PCTLUNCH	-0.000792	PERCENT IN LUNCH PROGRAM (LINEAR)
25 %LUNCH M	-0.003061	PERCENT IN LUNCH PROGRAM - MISSING
26 PCTWHT1	0.108255	PERCENT WHITE IN SCHOOL (0-49 %) WHT MIN
27 PCTWHT2	0.009537	PERCENT WHITE IN SCHOOL (50-79%) INTGRAT
28 AGE/GRD2	-0.217713	AGE X GRADE: MODAL AGE, < MODAL GRADE
29 AGE/GRD3	0.365221	AGE X GRADE: MODAL AGE, MODAL GRADE,MISS
30 AGE/GRD4	1.347927	AGE X GRADE: MODAL AGE, > MODAL GRADE
31 AGE/GRD5	0.156328	AGE X GRADE: > MODAL AGE, MODAL GRADE
32 SCHTYPE	-0.014791	SCHOOL TYPE: NON-PUBLIC
33 HW HELP	-0.127049	SOMEONE AT HOME HELPS WITH HW:>ONCE WEEK
34 PRESCH_Y	0.169416	WENT TO PRESCHOOL? (YES)
35 SINGLEP1	0.084494	HOW MANY PARENTS AT HOME? (BOTH)
36 MOMHOME1	0.260594	DOES YOUR MOTHER LIVE AT HOME? (YES)
37 MOMWORKY	-0.106985	DOES YOUR MOTHER WORK FOR PAY? (YES)
38 PGS RD14	0.119292	PAGES/DAY READ FOR SCHOOL (6 OR MORE)
39 PGS RD13	0.022448	PAGES/DAY READ FOR SCHOOL (11 OR MORE)
40 SCHMATH	-0.008494	SCHOOL LEVEL PROFICIENCY - MATH
41 SCHMATHM	0.046223	SCHOOL LEVEL PROFICIENCY MATH (MISSING)
42 SCHREAD	0.268947	SCHOOL LEVEL PROFICIENCY - READING
43 SCHREADM	-0.475073	SCHOOL LEVEL PROFICIENCY READING (MISSG)
44 SCHSCI	0.038513	SCHOOL LEVEL PROFICIENCY - SCIENCE
45 SCHSCIM	0.520162	SCHOOL LEVEL PROFICIENCY SCIENCE (MISSG)
46 SEASON_W	0.004893	SEASON - WINTER
47 MAGAZINE	0.021893	DO YOU GET ANY MAGS. FOR YOURSELF (LIN)
48 OWNBOOKS	0.059338	DO YOU OWN ANY BOOKS (NON SCHOOL) (LIN)
49 READTO	-0.009831	HOW OFTEN ARE YOU READ TO AT HOME (LIN)

Table F-11 (continued)
Estimated Effects for Reading Answer Booklet Bridge Conditioning Variable Contrasts, Age 9/Grade 4

Contrast	Estimated Effect	Description
50 READLOUD	0.303869	HOW OFTEN DO YOU READ OUT LOUD (LIN)
51 FOR_FUN	-0.041815	HOW OFTEN DO YOU READ FOR FUN (LIN)
52 TELLFRND	-0.087074	HOW OFTEN TELL A FRIEND ABOUT A BOOK
53 LIBRARY	0.420217	HOW OFTEN USE LIBRARY (LIN)
54 SPEND_$	-0.064971	HOW OFTEN SPEND YOUR MONEY ON BOOK (LIN)
55 READER	0.491689	WHAT KIND OF READER ARE YOU (LIN)
56 CHOOSEBK	-0.084020	READING TEACHER ASK YOU TO CHOOSE BOOK

Table F-12
Estimated Effects for Reading Trend Conditioning Variable Contrasts, Age 9

Contrast	Estimated Effect	Description
1 OVERALL	-1.044661	OVERALL CONSTANT
2 GENDER2	0.207947	SEX (FEMALE)
3 ETHNIC-2	-0.491551	ETHNICITY (BLACK)
4 ETHNIC-3	-0.323793	ETHNICITY (HISPANIC)
5 ETHNIC-4	-0.275318	ETHNICITY (ASIAN)
6 STOC-2	0.253260	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7 STOC-3	0.205837	SIZE AND TYPE OF COMMUNITY (ALL OTHERS &
8 REGION-2	-0.252109	REGION (SOUTHEAST)
9 REGION-3	-0.125535	REGION (CENTRAL)
10 REGION-4	-0.092016	REGION (WEST)
11 PARED-2	0.316885	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12 PARED-3	0.306620	PARENTS EDUCATION (POST-HIGH SCHOOL)
13 PARED-4	0.364517	PARENTS EDUCATION (COLLEGE GRAD)
14 PARED_	0.180067	PARENTS EDUCATION (MISSING AND I DON'T K
15 HOMEITM	0.113929	# ITEMS IN HOME.
16 HOMEITM_	-0.765748	ITEMS IN HOME (MISSING)
17 TV	-0.066646	# HOURS OF TV WATCHING
18 TV_	-0.237307	TV WATCH (MISSING)
19 HOMEWORK	-0.004169	HOMEWORK
20 HOMEWK_	-0.420233	HOMEWORK (MISSING)
21 HLM-23	-0.273315	HOME LANGUAGE MINORITY (SPANISH , OTHER)
22 HLM-BLK	0.027046	HOME LANGUAGE MINORITY (MISSING)
23 PREAD1-4	0.182554	PAGES READ (MORE THAN 6)
24 PREAD-B	0.003146	PAGES READ (MISSING)
25 %LUNCH	-0.260819	PERCENT LUNCH (0-100%)
26 %LUNCH_	-0.181088	PERCENT LUCH (MISSING)
27 PWHITE49	-0.006772	PERCENT WHITE (0-49)
28 PWHITE79	-0.007329	PERCENT WHITE (50-79)
29 PWHIT100	0.045283	PERCENT WHITE (80-100)
30 COURSES	0.005216	COURSES TAKESN
31 COURSES_	-0.176601	COURSES TAKESN
32 MODAGE-2	-0.677695	MODAL AGE,< MODAL GRADE
33 MODAGE-3	-0.043521	MODAL AGE, MODAL GRADE; MISSING
34 MODAGE-4	0.259293	MODAL AGE, > MODAL GRADE
35 MODAGE-5	-0.300943	> MODAL AGE, MODAL GRADE

Table F-13
Estimated Effects for Reading Trend Conditioning Variable Contrasts, Age 13

Contrast	Estimated Effect	Description
1 OVERALL	-0.003681	OVERALL CONSTANT
2 GENDER2	0.268898	SEX (FEMALE)
3 ETHNIC-2	-0.304269	ETHNICITY (BLACK)
4 ETHNIC-3	-0.168526	ETHNICITY (HISPANIC)
5 ETHNIC-4	-0.031977	ETHNICITY (ASIAN)
6 STOC-2	0.221806	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7 STOC-3	0.038778	SIZE AND TYPE OF COMMUNITY (ALL OTHERS &
8 REGION-2	0.046012	REGION (SOUTHEAST)
9 REGION-3	0.004675	REGION (CENTRAL)
10 REGION-4	0.082674	REGION (WEST)
11 PARED-2	0.030716	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12 PARED-3	0.364426	PARENTS EDUCATION (POST-HIGH SCHOOL)
13 PARED-4	0.288297	PARENTS EDUCATION (COLLEGE GRAD)
14 PARED_	-0.054217	PARENTS EDUCATION (MISSING AND I DON'T K
15 HOMEITM	0.110200	# ITEMS IN HOME.
16 HOMEITM_	0.300504	ITEMS IN HOME (MISSING)
17 TV	-0.065129	# HOURS OF TV WATCHING
18 TV_	-0.477439	TV WATCH (MISSING)
19 HOMEWORK	0.062328	HOMEWORK
20 HOMEWK_	-0.201913	HOMEWORK (MISSING)
21 HLM-23	-0.272108	HOME LANGUAGE MINORITY (SPANISH , OTHER)
22 HLM-BLK	-0.223249	HOME LANGUAGE MINORITY (MISSING)
23 PREAD1-4	0.192145	PAGES READ (MORE THAN 6)
24 PREAD-B	-0.081391	PAGES READ (MISSING)
25 %LUNCH	-0.010708	PERCENT LUNCH (0-100%)
26 %LUNCH_	0.032506	PERCENT LUCH (MISSING)
27 PWHITE49	0.030041	PERCENT WHITE (0-49)
28 PWHITE79	131176	PERCENT WHITE (50-79)
29 PWHIT100	.179477	PERCENT WHITE (80-100)
30 COURSES	0.038459	COURSES TAKESN
31 COURSES_	0.032815	COURSES TAKESN
32 MODAGE-2	-1.048760	MODAL AGE,< MODAL GRADE
33 MODAGE-3	-0.636542	MODAL AGE, MODAL GRADE; MISSING
34 MODAGE-4	0.173077	MODAL AGE, > MODAL GRADE
35 MODAGE-5	-0.901907	> MODAL AGE, MODAL GRADE

Table F-14
Estimated Effects for Reading Trend Conditioning Variable Contrasts, Age 17

Contrast	Estimated Effect	Description
1 OVERALL	0.781880	OVERALL CONSTANT
2 GENDER2	0.151576	SEX (FEMALE)
3 ETHNIC-2	-0.384247	ETHNICITY (BLACK)
4 ETHNIC-3	-0.095321	ETHNICITY (HISPANIC)
5 ETHNIC-4	0.003337	ETHNICITY (ASIAN)
6 STOC-2	0.050359	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7 STOC-3	-0.034364	SIZE AND TYPE OF COMMUNITY (ALL OTHERS
8 REGION-2	-0.123607	REGION (SOUTHEAST)
9 REGION-3	-0.084248	REGION (CENTRAL)
10 REGION-4	-0.113251	REGION (WEST)
11 PARED-2	0.127232	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12 PARED-3	0.266044	PARENTS EDUCATION (POST-HIGH SCHOOL)
13 PARED-4	0.345004	PARENTS EDUCATION (COLLEGE GRAD)
14 PARED_	-0.183837	PARENTS EDUCATION (MISSING AND I DON'T K
15 HOMEITM	0.060181	# ITEMS IN HOME.
16 HOMEITM_	0.372708	ITEMS IN HOME (MISSING)
17 TV	-0.081278	# HOURS OF TV WATCHING
18 TV_	-0.363956	TV WATCH (MISSING)
19 HOMEWORK	0.134986	HOMEWORK
20 HOMEWK_	-0.805181	HOMEWORK (MISSING)
21 HLM-23	-0.336985	HOME LANGUAGE MINORITY (SPANISH , OTHER)
22 HLM-BLK	-0.263394	HOME LANGUAGE MINORITY (MISSING)
23 PREAD1-4	0.211470	PAGES READ (MORE THAN 6)
24 PREAD-B	-0.050261	PAGES READ (MISSING)
25 %LUNCH	-0.414709	PERCENT LUNCH (0-100%)
26 %LUNCH_	-0.027250	PERCENT LUCH (MISSING)
27 PWHITE49	-0.028151	PERCENT WHITE (0-49)
28 PWHITE79	0.048558	PERCENT WHITE (50-79)
29 PWHIT100		PERCENT WHITE (80-100)
30 MODAGE-2	-0.679503	MODAL AGE,< MODAL GRADE
31 MODAGE-3	-0.063743	MODAL AGE, MODAL GRADE; MISSING
32 MODAGE-4	0.193545	MODAL AGE, > MODAL GRADE
33 MODAGE-5	-0.477520	> MODAL AGE, MODAL GRADE

Table F-15
Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Mathematics Cross-sectional Conditioning Variables, Age 9/Grade 4

Contrast	Proportion of Variance	Description
2 GENDER2	0.94078	SEX (FEMALE)
3 ETHNIC2	0.86593	DERIVED RACE (BLACK)
4 ETHNIC3	0.93315	DERIVED RACE (HISPANIC)
5 ETHNIC4	0.88539	DERIVED RACE (ASIAN)
6 STOC3	0.88160	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7 STOC147	0.88815	SIZE AND TYPE OF COMMUNITY (NON-HI+LOW)
8 REGION2	0.90081	REGION (SOUTHEAST)
9 REGION3	0.94571	REGION (CENTRAL)
10 REGION4	0.92816	REGION (WEST)
11 PARED2	0.94183	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12 PARED3	0.95172	PARENTS EDUCATION (POST HIGH SCHOOL)
13 PARED4	0.93282	PARENTS EDUCATION (COLLEGE GRAD)
14 PARED_	0.91602	PARENTS EDUCATION (MISSING,I DON'T KNOW)
15 HOMEITM2	0.82775	3 ITEMS IN THE HOME
16 HOMEITM3	0.78930	4 ITEMS IN THE HOME
17 TVWATCHL	0.98866	# HOURS OF DAILY TV VIEWING (LINEAR)
18 TVWATCHQ	0.98930	# HOURS OF DAILY TV VIEWING (QUADRATIC)
19 LANGHM23	0.94611	LANG OTHER THAN ENG AT HOME(SOME/ALWAYS)
20 HW-NO	0.88687	HOMEWORK (NONE ASSIGNED)
21 HW-YES	0.88951	HOMEWORK (YES, ASSIGNED)
22 HWLINEAR	0.94510	HOMEWORK (LINEAR)
23 HW:QUAD	0.86123	HOMEWORK (QUADRATIC)
24 PCTLUNCH	0.80430	PERCENT IN LUNCH PROGRAM (LINEAR)
25 %LUNCH M	0.89841	PERCENT IN LUNCH PROGRAM - MISSING
26 PCTWHT1	0.79330	PERCENT WHITE IN SCHOOL (0-49 %) WHT MIN
27 PCTWHT2	0.91452	PERCENT WHITE IN SCHOOL (50-79%) INTGRAT
28 AGE/GRD2	0.79818	AGE X GRADE: MODAL AGE, < MODAL GRADE
29 AGE/GRD3	0.93880	AGE X GRADE: MODAL AGE, MODAL GRADE,MISS
30 AGE/GRD4	0.99354	AGE X GRADE: MODAL AGE, > MODAL GRADE
31 AGE/GRD5	0.95594	AGE X GRADE: > MODAL AGE, MODAL GRADE
32 SCHTYPE	0.78734	SCHOOL TYPE: NON-PUBLIC
33 HW HELP	0.91766	SOMEONE AT HOME HELPS WITH HW:>ONCE WEEK
34 PRESCHOL	0.96987	WENT TO PRESCHOOL - YES
35 SINGLEP1	0.80722	HOW MANY PARENTS AT HOME? (BOTH)
36 MOMHOME1	0.80207	DOES YOUR MOTHER LIVE AT HOME? (YES)
37 MOMWORKY	0.98349	DOES YOUR MOTHER WORK FOR PAY? (YES)
38 PGS RD14	0.82044	PAGES/DAY READ FOR SCHOOL (6 OR MORE)
39 PGS RD13	0.82426	PAGES/DAY READ FOR SCHOOL (11 OR MORE)
40 SCHMATH	0.76754	SCHOOL LEVEL PROFICIENCY - MATH
41 SCHMATHM	0.71882	SCHOOL LEVEL PROFICIENCY MATH (MISSING)
42 SCHREAD	0.78943	SCHOOL LEVEL PROFICIENCY - READING
43 SCHREADM	0.91445	SCHOOL LEVEL PROFICIENCY READING (MISSG)
44 SCHSCI	0.79241	SCHOOL LEVEL PROFICIENCY - SCIENCE
45 SCHSCIM	0.95637	SCHOOL LEVEL PROFICIENCY SCIENCE (MISSG)
46 TEXTBOK1	0.93870	DO TEXTBOOK PROBLEMS (1-EVERY DAY)

Table F-15 (continued)
Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Mathematics Cross-sectional Conditioning Variables, Age 9/Grade 4

Contrast	Proportion of Variance	Description
47 TEXTBOK2	0.95862	DO TEXTBOOK PROBLEMS (2-SEVERAL/WEEK)
48 TEXTBOK3	0.97215	DO TEXTBOOK PROBLEMS (3-ONCE/WEEK)
49 TEXTBOK4	0.97617	DO TEXTBOOK PROBLEMS (4- <ONCE/WEEK)
50 WORKSHT1	0.96507	DO WORKSHEET PROBLEMS (1-EVERY DAY,MISS)
51 WORKSHT2	0.97564	DO WORKSHEET PROBLEMS (2-SEVERAL/WEEK)
52 WORKSHT3	0.97708	DO WORKSHEET PROBLEMS (3-ONCE/WEEK)
53 WORKSHT4	0.98323	DO WORKSHEET PROBLEMS (4-<ONCE/WEEK)
54 SMLLGRP1	0.88675	WORK IN SMALL GROUPS (1-EVERY DAY,MISS)
55 SMLLGRP2	0.92169	WORK IN SMALL GROUPS (2-SEVERAL/WEEK)
56 SMLLGRP3	0.89629	WORK IN SMALL GROUPS (3-ONCE/WEEK)
57 SMLLGRP4	0.88673	WORK IN SMALL GROUPS (4-<ONCE/WEEK)
58 W/OBJEC1	0.92337	WORK W/ SMALL OBJECTS(1-EVERY DAY,MISS)
59 W/OBJEC2	0.93236	WORK W/ SMALL OBJECTS(2-SEVERAL/WEEK)
60 W/OBJEC3	0.91423	WORK W/ SMALL OBJECTS(3-ONCE/WEEK)
61 W/OBJEC4	0.88764	WORK W/ SMALL OBJECTS(4-<ONCE/WEEK)
62 USECALC1	0.94172	USES CALCULATOR (1-EVERY DAY)
63 USECALC2	0.92051	USES CALCULATOR (2-SEVERAL/WEEK)
64 USECALC3	0.87413	USES CALCULATOR (3-ONCE/WEEK)
65 USECALC4	0.84751	USES CALCULATOR (4-<ONCE/WEEK)
66 USECOMP1	0.88709	USES COMPUTER (1-EVERY DAY,MISSING)
67 USECOMP2	0.91086	USES COMPUTER (2-SEVERAL/WEEK)
68 USECOMP3	0.87112	USES COMPUTER (3-ONCE/WEEK)
69 USECOMP4	0.89232	USES COMPUTER (4-<ONCE/WEEK)
70 TAKTEST1	0.98340	TAKES TESTS (1-EVERY DAY,MISSING)
71 TAKTEST2	0.98761	TAKES TESTS (2-SEVERAL/WEEK)
72 TAKTEST3	0.97278	TAKES TESTS (3-ONCE/WEEK)
73 TAKTEST4	0.97051	TAKES TESTS (4-<ONCE/WEEK)
74 TCHEXPL1	0.91012	TEACHER EXPLAINS CALCULATOR USE (YES)
75 TCHEXPL2	0.90905	TEACHER EXPLAINS CALCULATOR USE (NO)
76 HW DONE1	0.97499	AMOUNT OF HOMEWORK DONE (NONE)
77 HWDONE23	0.96087	AMOUNT OF HOMEWORK DONE (15-30 MINUTES)
78 HW DONE4	0.98761	AMOUNT OF HOMEWORK DONE (45 MINUTES)
79 HWDONE56	0.96315	AMOUNT OF HOMEWORK DONE (60, > HOUR)
80 LIKMATH2	0.82920	LIKES MATHEMATICS (UNDECIDED)
81 LIKMATH1	0.82158	LIKES MATHEMATICS (AGREE)
82 USEMATH2	0.84096	USE MATHEMATICS (UNDECIDED)
83 USEMATH1	0.84348	USE MATHEMATICS (AGREE)
84 GOODMAT2	0.86258	GOOD AT MATHEMATICS (UNDECIDED)
85 GOODMAT1	0.86144	GOOD AT MATHEMATICS (AGREE)
86 MATHBOY1	0.97444	MATHEMATICS IS FOR BOYS(AGREE)
87 MATHBOY2	0.98964	MATHEMATICS IS FOR BOYS(UNDECIDED)
88 USEFUL_2	0.83718	USEFUL FOR EVERYDAY PROBLEMS (UNDECIDED)
89 USEFUL_1	0.82519	USEFUL FOR EVERYDAY PROBLEMS (AGREE)
90 TGEOMTY1	0.94695	TEACHER: GEOMETRY COURSES TAKEN(NONE,OM)
91 TGEOMTY2	0.95467	TEACHER: GEOMETRY COURSES TAKEN(ONE)

Table F-15 (continued)
Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Mathematics Cross-sectional Conditioning Variables, Age 9/Grade 4

Contrast	Proportion of Variance	Description
92 TGEOMT34	0.76421	TEACHER: GEOMETRY COURSES TAKEN(2 OR 3)
93 TABSALG1	0.92757	TEACHER: ABSTRACT ALGEBRA COURSES (0,OM)
94 TABSALG2	0.86174	TEACHER: ABSTRACT ALGEBA COURSES (ONE)
95 TABSAL34	0.78518	TEACHER: ABSTRACT ALGEBRA COURSES (2,3)
96 TCALCUL1	0.93721	TEACHER: CALCULUS COURSES TAKEN (0,OMIT)
97 TCALCU23	0.77457	TEACHER: CALCULUS COURSES TAKEN (1,2)
98 TCALCUL4	0.82257	TEACHER: CALCULUS COURSES TAKEN (THREE)
99 TRESOUR1	0.98261	TEACHER: AMNT OF RESOURCES (ALL I NEED)
100 TRESOUR2	0.99378	TEACHER: AMNT OF RESOURCES(MOST OF NEED)
101 TRESOU34	0.98637	TEACHER: AMNT OF RESOURCES (SOME/NO RES)
102 TABILTY1	0.94486	TEACHER: ABILITY OF CLASS (HIGH)
103 TABILT24	0.98645	TEACHER: ABILITY OF CLASS (AVG,MIXED,OM)
104 TABILTY3	0.98271	TEACHER: ABILITY OF CLASS (LOW)
105 TMATHHWL	0.79126	TEACHER: AMNT MATH HOMEWORK DONE -LINEAR
106 TMATHHWM	0.98689	TEACHER: MATH HOMEWORK DONE(MATCHED)
107 TTEXTBK1	0.96490	TEACHER: DO TEXTBOOK PROBLEMS(EVERY DAY)
108 TTEXTBK2	0.98020	TEACHER: DO TEXTBOOK PROBLEMS(SEVERAL/W)
109 TTXTBK35	0.91847	TEACHER: DO TEXTBOOK PROBS(1/WK OR <,OM)
110 TWKSHT12	0.98994	TEACHER: DO WORKSHEET PROBS(DAILY,SEV/W)
111 TWKSHET3	0.98623	TEACHER: DO WORKSHEET PROBLEMS (1/WEEK)
112 TWKSHT45	0.98419	TEACHER: DO WORKSHEET PROBLEM (<1/WEEK)
113 TSMGRP12	0.97449	TEACHER: WORK IN SMALL GROUP(DAILY,SEV/W
114 TSMLGRP3	0.97515	TEACHER: WORK IN SMALL GROUPS (1/WEEK)
115 TSMGRP45	0.98193	TEACHER: WORK IN SMALL GROUPS (<1/WEEK)
116 TEMPMEAL	0.86662	TEACHER: EMPHASIS ON MEASURMENT - LINEAR
117 TEMPMEAM	0.98314	TEACHER: EMPHASIS ON MEASURMENT(MATCHED)
118 TEMPALGL	0.58071	TEACHER: EMPHASIS ON ALGEBRA - LINEAR
119 TEMPALGM	0.96892	TEACHER: EMPHASIS ON ALGEBRA (MATCHED)
120 TEMPCOML	0.83616	TEACHER: EMPHASIS-COMMUNICATING (LINEAR)
121 TEMPCOMM	0.97917	TEACHER: EMPHASIS-COMMUNICATING(MATCHED)
122 TCLCCLS1	0.82134	TEACHER: USE CALCULATORS IN CLASS (YES)
123 TCLCCLSM	0.99146	TEACHER: USE CALCULATORS/CLASS(MATCHED)
124 TCLCTST1	0.88423	TEACHER: USE CALCULATORS ON TEST (YES)
12[illegible] [illegible]LCTSTM	0.99146	TEACHER: USE CALCULATOR ON TEST(MATCHED)
126 TMATCH_P	0.95833	TEACHER: MATCH STATUS (PARTIAL)
127 TMATC[illegible]U	0.98943	TEACHER: MATCH STATUS (UNMATCHED)
128 TMTCHXR1	0.87644	TEACHER: MATCH STATUS X RACE 1
129 TMTCHXR2	0.89471	TEACHER: MATCH STATUS X RACE 2
130 TMTCHXR3	0.90300	TEACHER: MATCH STATUS X RACE 3
131 TMTCHXR4	0.91552	TEACHER: MATCH STATUS X RACE 4
132 TMTCHXR5	0.93427	TEACHER: MATCH STATUS X RACE 5
133 TMTCHXR6	0.93902	TEACHER: MATCH STATUS X RACE 6
134 TMTCHXR7	0.86099	TEACHER: MATCH STATUS X RACE 7
135 TMTCHXR8	0.90895	TEACHER: MATCH STATUS X RACE 8
136 TMTCHXP1	0.91068	TEACHER: MATCH STATUS X PARED 1

Table F-15 (continued)
Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Mathematics Cross-sectional Conditioning Variables, Age 9/Grade 4

Contrast	Proportion of Variance	Description
137 TMTCHXP2	0.91214	TEACHER: MATCH STATUS X PARED 2
138 TMTCHXP3	0.92722	TEACHER: MATCH STATUS X PARED 3
139 TMTCHXP4	0.93433	TEACHER: MATCH STATUS X PARED 4
140 TMTCHXP5	0.89761	TEACHER: MATCH STATUS X PARED 5
141 TMTCHXP6	0.89426	TEACHER: MATCH STATUS X PARED 6
142 TMTCHXP7	0.77508	TEACHER: MATCH STATUS X PARED 7
143 TMTCHXP8	0.79179	TEACHER: MATCH STATUS X PARED 8
144 TMTCHXR1	0.83235	TEACHER: MATCH STATUS X REGION 1
145 TMTCHXR2	0.91556	TEACHER: MATCH STATUS X REGION 2
146 TMTCHXR3	0.93588	TEACHER: MATCH STATUS X REGION 3
147 TMTCHXR4	0.93367	TEACHER: MATCH STATUS X REGION 4
148 TMTCHXR5	0.95667	TEACHER: MATCH STATUS X REGION 5
149 TMTCHXR6	0.95613	TEACHER: MATCH STATUS X REGION 6
150 TMTCHXS1	0.95689	TEACHER: MATCH STATUS X SEX 1
151 TMTCHXS2	0.95232	TEACHER: MATCH STATUS X SEX 2
152 SEASON_W	0.88326	SEASON OF ASSESSMENT (WINTER)

Table F-16
Estimated Effects for the Mathematics Cross-sectional Principal Components
Age 9/Grade 4

Variable	Estimated Effect	Description
1 OVERALL	-0.912903	OVERALL CONSTANT
2 PCFS1	0.012600	PRINCIPAL COMPONENT FACTOR SCORE 1
3 PCFS2	-0.067029	PRINCIPAL COMPONENT FACTOR SCORE 2
4 PCFS3	0.020968	PRINCIPAL COMPONENT FACTOR SCORE 3
5 PCFS4	-0.016482	PRINCIPAL COMPONENT FACTOR SCORE 4
6 PCFS5	0.029434	PRINCIPAL COMPONENT FACTOR SCORE 5
7 PCFS6	-0.016166	PRINCIPAL COMPONENT FACTOR SCORE 6
8 PCFS7	0.016446	PRINCIPAL COMPONENT FACTOR SCORE 7
9 PCFS8	-0.010461	PRINCIPAL COMPONENT FACTOR SCORE 8
10 PCFS9	0.039742	PRINCIPAL COMPONENT FACTOR SCORE 9
11 PCFS10	-0.065709	PRINCIPAL COMPONENT FACTOR SCORE 10
12 PCFS11	-0.017840	PRINCIPAL COMPONENT FACTOR SCORE 11
13 PCFS12	0.014821	PRINCIPAL COMPONENT FACTOR SCORE 12
14 PCFS13	-0.001605	PRINCIPAL COMPONENT FACTOR SCORE 13
15 PCFS14	-0.004535	PRINCIPAL COMPONENT FACTOR SCORE 14
16 PCFS15	-0.025958	PRINCIPAL COMPONENT FACTOR SCORE 15
17 PCFS16	0.017357	PRINCIPAL COMPONENT FACTOR SCORE 16
18 PCFS17	-0.027237	PRINCIPAL COMPONENT FACTOR SCORE 17
19 PCFS18	0.002086	PRINCIPAL COMPONENT FACTOR SCORE 18
20 PCFS19	0.011380	PRINCIPAL COMPONENT FACTOR SCORE 19
21 PCFS20	0.030469	PRINCIPAL COMPONENT FACTOR SCORE 20
22 PCFS21	0.011816	PRINCIPAL COMPONENT FACTOR SCORE 21
23 PCFS22	-0.004663	PRINCIPAL COMPONENT FACTOR SCORE 22
24 PCFS23	-0.004253	PRINCIPAL COMPONENT FACTOR SCORE 23
25 PCFS24	-0.015648	PRINCIPAL COMPONENT FACTOR SCORE 24
26 PCFS25	-0.026330	PRINCIPAL COMPONENT FACTOR SCORE 25
27 PCFS26	0.012853	PRINCIPAL COMPONENT FACTOR SCORE 26
28 PCFS27	0.000856	PRINCIPAL COMPONENT FACTOR SCORE 27
29 PCFS28	0.004160	PRINCIPAL COMPONENT FACTOR SCORE 28
30 PCFS29	-0.009209	PRINCIPAL COMPONENT FACTOR SCORE 29
31 PCFS30	0.001581	PRINCIPAL COMPONENT FACTOR SCORE 30
32 PCFS31	-0.016938	PRINCIPAL COMPONENT FACTOR SCORE 31
33 PCFS32	-0.008368	PRINCIPAL COMPONENT FACTOR SCORE 32
34 PCFS33	-0.005147	PRINCIPAL COMPONENT FACTOR SCORE 33
35 PCFS34	-0.009579	PRINCIPAL COMPONENT FACTOR SCORE 34
36 PCFS35	0.020254	PRINCIPAL COMPONENT FACTOR SCORE 35
37 PCFS36	-0.008536	PRINCIPAL COMPONENT FACTOR SCORE 36
38 PCFS37	-0.010084	PRINCIPAL COMPONENT FACTOR SCORE 37
39 PCFS38	-0.026907	PRINCIPAL COMPONENT FACTOR SCORE 38
40 PCFS39	-0.015423	PRINCIPAL COMPONENT FACTOR SCORE 39
41 PCFS40	0.015756	PRINCIPAL COMPONENT FACTOR SCORE 40
42 PCFS41	-0.009472	PRINCIPAL COMPONENT FACTOR SCORE 41
43 PCFS42	-0.007890	PRINCIPAL COMPONENT FACTOR SCORE 42
44 PCFS43	0.010351	PRINCIPAL COMPONENT FACTOR SCORE 43
45 PCFS44	0.000433	PRINCIPAL COMPONENT FACTOR SCORE 44
46 PCFS45	0.005615	PRINCIPAL COMPONENT FACTOR SCORE 45
47 PCFS46	-0.015231	PRINCIPAL COMPONENT FACTOR SCORE 46

Table F-16 (continued)
Estimated Effects for the Mathematics Cross-sectional Principal Components
Age 9/Grade 4

	Variable	Estimated Effect	Description
48	PCFS47	-0.023609	PRINCIPAL COMPONENT FACTOR SCORE 47
49	PCFS48	0.004342	PRINCIPAL COMPONENT FACTOR SCORE 48
50	PCFS49	0.004475	PRINCIPAL COMPONENT FACTOR SCORE 49
51	PCFS50	0.021479	PRINCIPAL COMPONENT FACTOR SCORE 50
52	PCFS51	-0.023332	PRINCIPAL COMPONENT FACTOR SCORE 51
53	PCFS52	-0.028768	PRINCIPAL COMPONENT FACTOR SCORE 52
54	PCFS53	0.008681	PRINCIPAL COMPONENT FACTOR SCORE 53
55	PCFS54	0.029242	PRINCIPAL COMPONENT FACTOR SCORE 54
56	PCFS55	-0.010916	PRINCIPAL COMPONENT FACTOR SCORE 55
57	PCFS56	-0.008574	PRINCIPAL COMPONENT FACTOR SCORE 56
58	PCFS57	0.024391	PRINCIPAL COMPONENT FACTOR SCORE 57
59	PCFS58	0.008857	PRINCIPAL COMPONENT FACTOR SCORE 58
60	PCFS59	0.019922	PRINCIPAL COMPONENT FACTOR SCORE 59
61	PCFS60	0.000587	PRINCIPAL COMPONENT FACTOR SCORE 60
62	PCFS61	0.003191	PRINCIPAL COMPONENT FACTOR SCORE 61
63	PCFS62	0.007364	PRINCIPAL COMPONENT FACTOR SCORE 62
64	PCFS63	0.029724	PRINCIPAL COMPONENT FACTOR SCORE 63
65	PCFS64	-0.005934	PRINCIPAL COMPONENT FACTOR SCORE 64
66	PCFS65	-0.006151	PRINCIPAL COMPONENT FACTOR SCORE 65
67	PCFS66	0.005063	PRINCIPAL COMPONENT FACTOR SCORE 66
68	PCFS67	0.033424	PRINCIPAL COMPONENT FACTOR SCORE 67
69	PCFS68	-0.042041	PRINCIPAL COMPONENT FACTOR SCORE 68
70	PCFS69	0.015255	PRINCIPAL COMPONENT FACTOR SCORE 69
71	PCFS70	0.004948	PRINCIPAL COMPONENT FACTOR SCORE 70
72	PCFS71	-0.017023	PRINCIPAL COMPONENT FACTOR SCORE 71
73	PCFS72	0.022202	PRINCIPAL COMPONENT FACTOR SCORE 72
74	PCFS73	-0.000135	PRINCIPAL COMPONENT FACTOR SCORE 73
75	PCFS74	0.027243	PRINCIPAL COMPONENT FACTOR SCORE 74
76	PCFS75	0.005557	PRINCIPAL COMPONENT FACTOR SCORE 75
77	PCFS76	-0.008495	PRINCIPAL COMPONENT FACTOR SCORE 76
78	PCFS77	-0.007614	PRINCIPAL COMPONENT FACTOR SCORE 77
79	PCFS78	0.000460	PRINCIPAL COMPONENT FACTOR SCORE 78
80	PCFS79	-0.072869	PRINCIPAL COMPONENT FACTOR SCORE 79
81	PCFS80	0.015915	PRINCIPAL COMPONENT FACTOR SCORE 80
82	PCFS81	0.058948	PRINCIPAL COMPONENT FACTOR SCORE 81
83	PCFS82	-0.037889	PRINCIPAL COMPONENT FACTOR SCORE 82
84	PCFS83	-0.007089	PRINCIPAL COMPONENT FACTOR SCORE 83
85	PCFS84	-0.022559	PRINCIPAL COMPONENT FACTOR SCORE 84

Table F-17
Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Mathematics Cross-sectional Conditioning Variables, Age 13/Grade 8

Contrast	Proportion of Variance	Description
2 GENDER2	0.90220	SEX (FEMALE)
3 ETHNIC2	0.87938	DERIVED RACE (BLACK)
4 ETHNIC3	0.81392	DERIVED RACE (HISPANIC)
5 ETHNIC4	0.87367	DERIVED RACE (ASIAN)
6 STOC3	0.85641	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7 STOC147	0.88332	SIZE AND TYPE OF COMMUNITY (NON-HI+LOW)
8 REGION2	0.91304	REGION (SOUTHEAST)
9 REGION3	0.90865	REGION (CENTRAL)
10 REGION4	0.91162	REGION (WEST)
11 PARED2	0.89415	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12 PARED3	0.89843	PARENTS EDUCATION (POST HIGH SCHOOL)
13 PARED4	0.88777	PARENTS EDUCATION (COLLEGE GRAD)
14 PARED_	0.93274	PARENTS EDUCATION (MISSING,I DON'T KNOW)
15 HOMEITM2	0.86980	3 ITEMS IN THE HOME
16 HOMEITM3	0.84237	4 ITEMS IN THE HOME
17 TVWATCHL	0.98519	# HOURS OF DAILY TV VIEWING (LINEAR)
18 TVWATCHQ	0.98493	# HOURS OF DAILY TV VIEWING (QUADRATIC)
19 LANGHM23	0.75981	LANG OTHER THAN ENG AT HOME(SOME/ALWAYS)
20 HW-NO	0.96075	HOMEWORK (NONE ASSIGNED)
21 HW-YES	0.96365	HOMEWORK (YES, ASSIGNED)
22 HWLINEAR	0.94800	HOMEWORK (LINEAR)
23 HW:QUAD	0.92689	HOMEWORK (QUADRATIC)
24 PCTLUNCH	0.84981	PERCENT IN LUNCH PROGRAM (LINEAR)
25 %LUNCH M	0.81933	PERCENT IN LUNCH PROGRAM - MISSING
26 PCTWHT1	0.82179	PERCENT WHITE IN SCHOOL (0-49 %) WHT MIN
27 PCTWHT2	0.86362	PERCENT WHITE IN SCHOOL (50-79%) INTGRAT
28 AGE/GRD2	0.83548	AGE X GRADE: MODAL AGE, < MODAL GRADE
29 AGE/GRD3	0.94682	AGE X GRADE: MODAL AGE, MODAL GRADE,MISS
30 AGE/GRD4	0.98708	AGE X GRADE: MODAL AGE, > MODAL GRADE
31 AGE/GRD5	0.95450	AGE X GRADE: > MODAL AGE, MODAL GRADE
32 SCHTYPE	0.88876	SCHOOL TYPE: NON-PUBLIC
33 HW HELP	0.97738	SOMEONE AT HOME HELPS WITH HW:>ONCE WEEK
34 SINGLEP1	0.78914	HOW MANY PARENTS AT HOME? (BOTH)
35 MOMHOME1	0.76595	DOES YOUR MOTHER LIVE AT HOME? (YES)
36 MOMWORKY	0.94033	DOES YOUR MOTHER WORK FOR PAY? (YES)
37 PGS RD14	0.79684	PAGES/DAY READ FOR SCHOOL (6 OR MORE)
38 PGS RD13	0.80644	PAGES/DAY READ FOR SCHOOL (11 OR MORE)
39 EXPHSGRD	0.59326	EXPECT TO GRADUATE FROM HIGH SCHOOL(YES)
40 SCHMSS12	0.97542	SCHOOL DAYS MISSED LAST MONTH (0-2 DAYS)
41 BEHAVRNM	0.80898	RULES OF BEHAVIOR ARE STRICT (NON-MISSG)
42 BEHAVR-L	0.83151	RULES OF BEHAVIOR ARE STRICT (LINEAR)
43 SAFETYNM	0.85875	DON'T FEEL SAFE AT SCHOOL (NON-MISSG)
44 SAFETY-L	0.60901	DON'T FEEL SAFE AT SCHOOL (LINEAR)
45 DISRPTNM	0.82055	STUDENTS OFTEN DISRUPT CLASS (NON-MISSG)
46 DISRPT-L	0.73772	STUDENTS OFTEN DISRUPT CLASS (LINEAR)

Table F-17 (continued)
Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Mathematics Cross-sectional Conditioning Variables, Age 13/Grade 8

Contrast	Proportion of Variance	Description
47 SCHMATH	0.80079	SCHOOL LEVEL PROFICIENCY - MATH (CONT)
48 SCHMATHM	0.73450	SCHOOL LEVEL PROFICIENCY MATH (MISSING)
49 SCHREAD	0.82820	SCHOOL LEVEL PROFICIENCY - READING(CONT)
50 SCHREADM	0.95481	SCHOOL LEVEL PROFICIENCY READING (MISSG)
51 SCHSCI	0.81057	SCHOOL LEVEL PROFICIENCY - SCIENCE(CONT)
52 SCHSCIM	0.95461	SCHOOL LEVEL PROFICIENCY SCIENCE (MISSG)
53 TEXTBOK1	0.96277	DO TEXTBOOK PROBLEMS (1-EVERY DAY)
54 TEXTBOK2	0.97284	DO TEXTBOOK PROBLEMS (2-SEVERAL/WEEK)
55 TEXTBOK3	0.98152	DO TEXTBOOK PROBLEMS (3-ONCE/WEEK)
56 TEXTBOK4	0.98261	DO TEXTBOOK PROBLEMS (4- <ONCE/WEEK)
57 WORKSHT1	0.94458	DO WORKSHEET PROBLEMS (1-EVERY DAY,MISS)
58 WORKSHT2	0.96109	DO WORKSHEET PROBLEMS (2-SEVERAL/WEEK)
59 WORKSHT3	0.95844	DO WORKSHEET PROBLEMS (3-ONCE/WEEK)
60 WORKSHT4	0.95194	DO WORKSHEET PROBLEMS (4-<ONCE/WEEK)
61 SMLLGRP1	0.90378	WORK IN SMALL GROUPS (1-EVERY DAY,MISS)
62 SMLLGRP2	0.93358	WORK IN SMALL GROUPS (2-SEVERAL/WEEK)
63 SMLLGRP3	0.89745	WORK IN SMALL GROUPS (3-ONCE/WEEK)
64 SMLLGRP4	0.86700	WORK IN SMALL GROUPS (4-<ONCE/WEEK)
65 W/OBJEC1	0.92693	WORK W/ SMALL OBJECTS(1-EVERY DAY,MISS)
66 W/OBJEC2	0.91100	WORK W/ SMALL OBJECTS(2-SEVERAL/WEEK)
67 W/OBJEC3	0.90412	WORK W/ SMALL OBJECTS(3-ONCE/WEEK)
68 W/OBJEC4	0.88003	WORK W/ SMALL OBJECTS(4-<ONCE/WEEK)
69 USECALC1	0.89540	USES CALCULATOR (1-EVERY DAY)
70 USECALC2	0.93211	USES CALCULATOR (2-SEVERAL/WEEK)
71 USECALC3	0.93688	USES CALCULATOR (3-ONCE/WEEK)
72 USECALC4	0.92056	USES CALCULATOR (4-<ONCE/WEEK)
73 USECOMP1	0.91341	USES COMPUTER (1-EVERY DAY,MISSING)
74 USECOMP2	0.93405	USES COMPUTER (2-SEVERAL/WEEK)
75 USECOMP3	0.92500	USES COMPUTER (3-ONCE/WEEK)
76 USECOMP4	0.89598	USES COMPUTER (4-<ONCE/WEEK)
77 TAKTEST1	0.87692	TAKES TESTS (1-EVERY DAY,MISSING)
78 TAKTEST2	0.99429	TAKES TESTS (2-SEVERAL/WEEK)
79 TAKTEST3	0.99009	TAKES TESTS (3-ONCE/WEEK)
80 TAKTEST4	0.96733	TAKES TESTS (4-<ONCE/WEEK)
81 WRITREP1	0.83253	WRITES REPORTS (1-EVERY DAY,MISSING)
82 WRITREP2	0.93496	WRITES REPORTS (2-SEVERAL/WEEK)
83 WRITREP3	0.92684	WRITES REPORTS (3-ONCE/WEEK)
84 WRITREP4	0.89120	WRITES REPORTS (4-<ONCE/WEEK)
85 TCHEXPL1	0.96955	TEACHER EXPLAINS CALCULATOR USE (YES)
86 TCHEXPL2	0.96922	TEACHER EXPLAINS CALCULATOR USE (NO)
87 CALCCLS1	0.90284	USE CALCULATOR:CLASS PROBLMS(ALWAYS,MIS)
88 CALCCLS2	0.85508	USE CALCULATOR:CLASS PROBLEMS(SOMETIMES)
89 CALCHOM1	0.88304	USE CALCULATOR:HOME PROBLEMS(ALWAYS,MIS)
90 CALCHOM2	0.87382	USE CALCULATOR:HOME PROBLEMS (SOMETIMES)
91 CALCTST1	0.84865	USE CALCULATOR:ON TESTS (ALWAYS,MISSING)

Table F-17 (continued)
Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Mathematics Cross-sectional Conditioning Variables, Age 13/Grade 8

Contrast	Proportion of Variance	Description
92 CALCTST2	0.89049	USE CALCULATOR:ON TESTS (SOMETIMES)
93 MATHCLS2	0.89333	KIND OF MATH CLASS (8TH GRADE MATH)
94 MATHCLS3	0.96184	KIND OF MATH CLASS (PRE-ALGEBRA)
95 MATHCLS4	0.86503	KIND OF MATH CLASS (ALGEBRA)
96 MATHCLS5	0.92295	KIND OF MATH CLASS (OTHER)
97 HW DONE1	0.91218	AMOUNT OF HOMEWORK DONE (NONE)
98 HWDONE23	0.97771	AMOUNT OF HOMEWORK DONE (15-30 MINUTES)
99 HW DONE4	0.97486	AMOUNT OF HOMEWORK DONE (45 MINUTES)
100 HWDONE56	0.93229	AMOUNT OF HOMEWORK DONE (60, > HOUR)
101 LIKMATH4	0.90459	LIKES MATHEMATICS (DISAGREE)
102 LIKMATH3	0.92142	LIKES MATHEMATICS (UNDECIDED)
103 LIKMATH2	0.95188	LIKES MATHEMATICS (AGREE)
104 LIKMATH1	0.83451	LIKES MATHEMATICS (STRONGLY AGREE)
105 USEMATH4	0.97853	USE MATHEMATICS (DISAGREE)
106 USEMATH3	0.97799	USE MATHEMATICS (UNDECIDED)
107 USEMATH2	0.97813	USE MATHEMATICS (AGREE)
108 USEMATH1	0.96730	USE MATHEMATICS (STRONGLY AGREE)
109 GOODMAT4	0.91534	GOOD AT MATHEMATICS (DISAGREE)
110 GOODMAT3	0.92787	GOOD AT MATHEMATICS (UNDECIDED)
111 GOODMAT2	0.95681	GOOD AT MATHEMATICS (AGREE)
112 GOODMAT1	0.83131	GOOD AT MATHEMATICS (STRONGLY AGREE)
113 MATBOY12	0.98777	MATHEMATICS IS FOR BOYS(AGREE/STR AGREE)
114 MATBOY34	0.97219	MATHEMATICS IS FOR BOYS(UNDECIDE/DISAGR)
115 MATBOYS5	0.96769	MATHEMATICS IS FOR BOYS(STRONGLY DISAGR)
116 USEFUL_4	0.97255	USEFUL FOR EVERYDAY PROBLEMS (DISAGREE)
117 USEFUL_3	0.97392	USEFUL FOR EVERYDAY PROBLEMS (UNDECIDED)
118 USEFUL_2	0.96677	USEFUL FOR EVERYDAY PROBLEMS (AGREE)
119 USEFUL_1	0.95708	USEFUL FOR EVERYDAY PROBLEMS (STR AGREE)
120 TGEOMTY1	0.79387	TEACHER: GEOMETRY COURSES TAKEN(NONE,OM)
121 TGEOMTY2	0.97170	TEACHER: GEOMETRY COURSES TAKEN(ONE)
122 TGEOMT34	0.87655	TEACHER: GEOMETRY COURSES TAKEN(2 OR 3)
123 TABSALG1	0.85856	TEACHER: ABSTRACT ALGEBRA COURSES (0,OM)
124 TABSALG2	0.97571	TEACHER: ABSTRACT ALGEBA COURSES (ONE)
125 TABSAL34	0.89982	TEACHER: ABSTRACT ALGEBRA COURSES (2,3)
126 TCALCUL1	0.82492	TEACHER: CALCULUS COURSES TAKEN (0,OMIT)
127 TCALCU23	0.95996	TEACHER: CALCULUS COURSES TAKEN (1,2)
128 TCALCUL4	0.90055	TEACHER: CALCULUS COURSES TAKEN (THREE)
129 TRESOUR1	0.97496	TEACHER: AMNT OF RESOURCES (ALL I NEED)
130 TRESOUR2	0.99049	TEACHER: AMNT OF RESOURCES(MOST OF NEED)
131 TRESOU34	0.98522	TEACHER: AMNT OF RESOURCES (SOME/NO RES)
132 TABILTY1	0.88218	TEACHER: ABILITY OF CLASS (HIGH)
133 TABILT24	0.97053	TEACHER: ABILITY OF CLASS (AVG,MIXED,OM)
134 TABILTY3	0.96463	TEACHER: ABILITY OF CLASS (LOW)
135 TMATHHWL	0.82424	TEACHER: AMNT MATH HOMEWORK DONE -LINEAR
136 TMATHHWM	0.99042	TEACHER: MATH HOMEWORK DONE(MATCHED)

Table F-17 (continued)
Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Mathematics Cross-sectional Conditioning Variables, Age 13/Grade 8

Contrast	Proportion of Variance	Description
137 TTEXTBK1	0.93471	TEACHER: DO TI XTBOOK PROBLEMS(EVERY DAY)
138 TTEXTBK2	0.96181	TEACHER: DO TEXTBOOK PROBLEMS(SEVERAL/W)
139 TTXTBK35	0.84532	TEACHER: DO TEXTBOOK PROBS(1/WK OR <,OM)
140 TWKSHT12	0.91475	TEACHER: DO WORKSHEET PROBS(DAILY,SEV/W)
141 TWKSHET3	0.99258	TEACHER: DO WORKSHEET PROBLEMS (1/WEEK)
142 TWKSHT45	0.89798	TEACHER: DO WORKSHEET PROBLEM (<1/WEEK)
143 TCLCCLS1	0.77207	TEACHER: USE CALCULATORS IN CLASS (YES)
144 TCLCCLSM	0.99358	TEACHER: USE CALCULATORS/CLASS(MATCHED)
145 TCLCTST1	0.75111	TEACHER: USE CALCULATORS ON TEST (YES)
146 TCLCTSTM	0.99358	TEACHER: USE CALCULATOR ON TEST(MATCHED)
147 TEMPN&OL	0.92552	TEACHER: EMPHASIS ON NUMOPERATNS-LINEAR
148 TEMPMEAL	0.81519	TEACHER: EMPHASIS ON MEASURMENT - LINEAR
149 TEMPMEAM	0.99148	TEACHER: EMPHASIS ON MEASURMENT(MATCHED)
150 TEMPALGL	0.89045	TEACHER: EMPHASIS ON ALGEBRA - LINEAR
151 TEMPALGM	0.99358	TEACHER: EMPHASIS ON ALGEBRA (MATCHED)
152 TMATCH_P	0.96189	TEACHER: MATCH STATUS (PARTIAL)
153 TMATCH_U	0.98560	TEACHER: MATCH STATUS (UNMATCHED)
154 TMTCHXR1	0.86909	TEACHER: MATCH STATUS X RACE 1
155 TMTCHXR2	0.87694	TEACHER: MATCH STATUS X RACE 2
156 TMTCHXR3	0.91032	TEACHER: MATCH STATUS X RACE 3
157 TMTCHXR4	0.90072	TEACHER: MATCH STATUS X RACE 4
158 TMTCHXR5	0.87037	TEACHER: MATCH STATUS X RACE 5
159 TMTCHXR6	0.84676	TEACHER: MATCH STATUS X RACE 6
160 TMTCHXR7	0.84773	TEACHER: MATCH STATUS X RACE 7
161 TMTCHXR8	0.90301	TEACHER: MATCH STATUS X RACE 8
162 TMTCHXP1	0.78906	TEACHER: MATCH STATUS X PARED 1
163 TMTCHXP2	0.82015	TEACHER: MATCH STATUS X PARED 2
164 TMTCHXP3	0.84526	TEACHER: MATCH STATUS X PARED 3
165 TMTCHXP4	0.86416	TEACHER: MATCH STATUS X PARED 4
166 TMTCHXP5	0.79865	TEACHER: MATCH STATUS X PARED 5
167 TMTCHXP6	0.82573	TEACHER: MATCH STATUS X PARED 6
168 TMTCHXP7	0.90364	TEACHER: MATCH STATUS X PARED 7
169 TMTCHXP8	0.81947	TEACHER: MATCH STATUS X PARED 8
170 TMTCHXR1	0.90376	TEACHER: MATCH STATUS X REGION 1
171 TMTCHXR2	0.90081	TEACHER: MATCH STATUS X REGION 2
172 TMTCHXR3	0.93340	TEACHER: MATCH STATUS X REGION 3
173 TMTCHXR4	0.91722	TEACHER: MATCH STATUS X REGION 4
174 TMTCHXR5	0.87516	TEACHER: MATCH STATUS X REGION 5
175 TMTCHXR6	0.87921	TEACHER: MATCH STATUS X REGION 6
176 TMTCHXS1	0.92933	TEACHER: MATCH STATUS X SEX 1
177 TMTCHXS2	0.92135	TEACHER: MATCH STATUS X SEX 2
178 SEASON_W	0.91084	SEASON OF ASSESSMENT (WINTER)

Table F-18
Estimated Effects for the Mathematics Cross-sectional Principal Components
Age 13/Grade 8

	Variable	Estimated Effect	Description
1	OVERALL	0.211935	OVERALL CONSTANT
2	PCFS1	0.014049	PRINCIPAL COMPONENT FACTOR SCORE 1
3	PCFS2	-0.062869	PRINCIPAL COMPONENT FACTOR SCORE 2
4	PCFS3	0.030632	PRINCIPAL COMPONENT FACTOR SCORE 3
5	PCFS4	-0.021711	PRINCIPAL COMPONENT FACTOR SCORE 4
6	PCFS5	-0.025008	PRINCIPAL COMPONENT FACTOR SCORE 5
7	PCFS6	-0.003182	PRINCIPAL COMPONENT FACTOR SCORE 6
8	PCFS7	-0.007614	PRINCIPAL COMPONENT FACTOR SCORE 7
9	PCFS8	-0.003435	PRINCIPAL COMPONENT FACTOR SCORE 8
10	PCFS9	-0.015741	PRINCIPAL COMPONENT FACTOR SCORE 9
11	PCFS10	-0.058391	PRINCIPAL COMPONENT FACTOR SCORE 10
12	PCFS11	-0.041618	PRINCIPAL COMPONENT FACTOR SCORE 11
13	PCFS12	0.000973	PRINCIPAL COMPONENT FACTOR SCORE 12
14	PCFS13	0.000383	PRINCIPAL COMPONENT FACTOR SCORE 13
15	PCFS14	0.033899	PRINCIPAL COMPONENT FACTOR SCORE 14
16	PCFS15	0.013662	PRINCIPAL COMPONENT FACTOR SCORE 15
17	PCFS16	-0.022889	PRINCIPAL COMPONENT FACTOR SCORE 16
18	PCFS17	0.005970	PRINCIPAL COMPONENT FACTOR SCORE 17
19	PCFS18	-0.001968	PRINCIPAL COMPONENT FACTOR SCORE 18
20	PCFS19	-0.003385	PRINCIPAL COMPONENT FACTOR SCORE 19
21	PCFS20	-0.007629	PRINCIPAL COMPONENT FACTOR SCORE 20
22	PCFS21	0.017602	PRINCIPAL COMPONENT FACTOR SCORE 21
23	PCFS22	0.002338	PRINCIPAL COMPONENT FACTOR SCORE 22
24	PCFS23	0.026393	PRINCIPAL COMPONENT FACTOR SCORE 23
25	PCFS24	-0.049207	PRINCIPAL COMPONENT FACTOR SCORE 24
26	PCFS25	0.023890	PRINCIPAL COMPONENT FACTOR SCORE 25
27	PCFS26	0.009745	PRINCIPAL COMPONENT FACTOR SCORE 26
28	PCFS27	0.022151	PRINCIPAL COMPONENT FACTOR SCORE 27
29	PCFS28	0.008057	PRINCIPAL COMPONENT FACTOR SCORE 28
30	PCFS29	0.025530	PRINCIPAL COMPONENT FACTOR SCORE 29
31	PCFS30	-0.004139	PRINCIPAL COMPONENT FACTOR SCORE 30
32	PCFS31	-0.010178	PRINCIPAL COMPONENT FACTOR SCORE 31
33	PCFS32	-0.014313	PRINCIPAL COMPONENT FACTOR SCORE 32
34	PCFS33	0.032264	PRINCIPAL COMPONENT FACTOR SCORE 33
35	PCFS34	-0.004962	PRINCIPAL COMPONENT FACTOR SCORE 34
36	PCFS35	-0.012760	PRINCIPAL COMPONENT FACTOR SCORE 35
37	PCFS36	0.004446	PRINCIPAL COMPONENT FACTOR SCORE 36
38	PCFS37	-0.005299	PRINCIPAL COMPONENT FACTOR SCORE 37
39	PCFS38	0.007856	PRINCIPAL COMPONENT FACTOR SCORE 38
40	PCFS39	-0.033705	PRINCIPAL COMPONENT FACTOR SCORE 39
41	PCFS40	-0.000037	PRINCIPAL COMPONENT FACTOR SCORE 40
42	PCFS41	0.009410	PRINCIPAL COMPONENT FACTOR SCORE 41
43	PCFS42	-0.002984	PRINCIPAL COMPONENT FACTOR SCORE 42
44	PCFS43	-0.007056	PRINCIPAL COMPONENT FACTOR SCORE 43
45	PCFS44	-0.010097	PRINCIPAL COMPONENT FACTOR SCORE 44
46	PCFS45	0.008759	PRINCIPAL COMPONENT FACTOR SCORE 45
47	PCFS46	0.029636	PRINCIPAL COMPONENT FACTOR SCORE 46

Table F-18 (continued)
Estimated Effects for the Mathematics Cross-sectional Principal Components
Age 13/Grade 8

	Variable	Estimated Effect	Description
48	PCFS47	-0.014222	PRINCIPAL COMPONENT FACTOR SCORE 47
49	PCFS48	0.005502	PRINCIPAL COMPONENT FACTOR SCORE 48
50	PCFS49	0.012162	PRINCIPAL COMPONENT FACTOR SCORE 49
51	PCFS50	-0.003722	PRINCIPAL COMPONENT FACTOR SCORE 50
52	PCFS51	0.016502	PRINCIPAL COMPONENT FACTOR SCORE 51
53	PCFS52	-0.011508	PRINCIPAL COMPONENT FACTOR SCORE 52
54	PCFS53	0.027347	PRINCIPAL COMPONENT FACTOR SCORE 53
55	PCFS54	0.029653	PRINCIPAL COMPONENT FACTOR SCORE 54
56	PCFS55	0.004815	PRINCIPAL COMPONENT FACTOR SCORE 55
57	PCFS56	0.007339	PRINCIPAL COMPONENT FACTOR SCORE 56
58	PCFS57	0.001648	PRINCIPAL COMPONENT FACTOR SCORE 57
59	PCFS58	0.017363	PRINCIPAL COMPONENT FACTOR SCORE 58
60	PCFS59	0.008952	PRINCIPAL COMPONENT FACTOR SCORE 59
61	PCFS60	0.017347	PRINCIPAL COMPONENT FACTOR SCORE 60
62	PCFS61	0.018348	PRINCIPAL COMPONENT FACTOR SCORE 61
63	PCFS62	0.001427	PRINCIPAL COMPONENT FACTOR SCORE 62
64	PCFS63	0.010881	PRINCIPAL COMPONENT FACTOR SCORE 63
65	PCFS64	0.025335	PRINCIPAL COMPONENT FACTOR SCORE 64
66	PCFS65	-0.015984	PRINCIPAL COMPONENT FACTOR SCORE 65
67	PCFS66	0.019857	PRINCIPAL COMPONENT FACTOR SCORE 66
68	PCFS67	0.020069	PRINCIPAL COMPONENT FACTOR SCORE 67
69	PCFS68	-0.005578	PRINCIPAL COMPONENT FACTOR SCORE 68
70	PCFS69	-0.001280	PRINCIPAL COMPONENT FACTOR SCORE 69
71	PCFS70	-0.019838	PRINCIPAL COMPONENT FACTOR SCORE 70
72	PCFS71	0.002508	PRINCIPAL COMPONENT FACTOR SCORE 71
73	PCFS72	0.059531	PRINCIPAL COMPONENT FACTOR SCORE 72
74	PCFS73	-0.011844	PRINCIPAL COMPONENT FACTOR SCORE 73
75	PCFS74	-0.001358	PRINCIPAL COMPONENT FACTOR SCORE 74
76	PCFS75	0.022795	PRINCIPAL COMPONENT FACTOR SCORE 75
77	PCFS76	0.010779	PRINCIPAL COMPONENT FACTOR SCORE 76
78	PCFS77	-0.003904	PRINCIPAL COMPONENT FACTOR SCORE 77
79	PCFS78	0.002143	PRINCIPAL COMPONENT FACTOR SCORE 78
80	PCFS79	-0.018100	PRINCIPAL COMPONENT FACTOR SCORE 79
81	PCFS80	-0.001666	PRINCIPAL COMPONENT FACTOR SCORE 80
82	PCFS81	-0.004126	PRINCIPAL COMPONENT FACTOR SCORE 81
83	PCFS82	0.018404	PRINCIPAL COMPONENT FACTOR SCORE 82
84	PCFS83	0.006649	PRINCIPAL COMPONENT FACTOR SCORE 83
85	PCFS84	-0.006022	PRINCIPAL COMPONENT FACTOR SCORE 84
86	PCFS85	-0.017920	PRINCIPAL COMPONENT FACTOR SCORE 85
87	PCFS86	-0.024725	PRINCIPAL COMPONENT FACTOR SCORE 86
88	PCFS87	-0.007297	PRINCIPAL COMPONENT FACTOR SCORE 87
89	PCFS88	-0.002668	PRINCIPAL COMPONENT FACTOR SCORE 88
90	PCFS89	-0.014320	PRINCIPAL COMPONENT FACTOR SCORE 89
91	PCFS90	0.017292	PRINCIPAL COMPONENT FACTOR SCORE 90
92	PCFS91	-0.022240	PRINCIPAL COMPONENT FACTOR SCORE 91
93	PCFS92	-0.032896	PRINCIPAL COMPONENT FACTOR SCORE 92
94	PCFS93	0.018833	PRINCIPAL COMPONENT FACTOR SCORE 93

Table F-18 (continued)
Estimated Effects for the Mathematics Cross-sectional Principal Components
Age 13/Grade 8

	Variable	Estimated Effect	Description
95	PCFS94	-0.023475	PRINCIPAL COMPONENT FACTOR SCORE 94
96	PCFS95	0.009489	PRINCIPAL COMPONENT FACTOR SCORE 95
97	PCFS96	0.001374	PRINCIPAL COMPONENT FACTOR SCORE 96
98	PCFS97	-0.014128	PRINCIPAL COMPONENT FACTOR SCORE 97
99	PCFS98	0.000402	PRINCIPAL COMPONENT FACTOR SCORE 98
100	PCFS99	-0.001748	PRINCIPAL COMPONENT FACTOR SCORE 99
101	PCFS100	-0.005336	PRINCIPAL COMPONENT FACTOR SCORE 100

Table F-19
Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Mathematics Cross-sectional Conditioning Variables, Age 17/Grade 12

Contrast	Proportion of Variance	Description
2 GENDER2	0.76365	SEX (FEMALE)
3 ETHNIC2	0.71782	DERIVED RACE (BLACK)
4 ETHNIC3	0.83332	DERIVED RACE (HISPANIC)
5 ETHNIC4	0.92380	DERIVED RACE (ASIAN)
6 STOC3	0.85935	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7 STOC147	0.84564	SIZE AND TYPE OF COMMUNITY (NON-HI+LOW)
8 REGION2	0.90120	REGION (SOUTHEAST)
9 REGION3	0.89506	REGION (CENTRAL)
10 REGION4	0.89136	REGION (WEST)
11 PARED2	0.96253	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12 PARED3	0.96554	PARENTS EDUCATION (POST HIGH SCHOOL)
13 PARED4	0.95446	PARENTS EDUCATION (COLLEGE GRAD)
14 PARED_	0.98251	PARENTS EDUCATION (MISSING,I DON'T KNOW)
15 HOMEITM2	0.88297	3 ITEMS IN THE HOME
16 HOMEITM3	0.86405	4 ITEMS IN THE HOME
17 TVWATCHL	0.97884	# HOURS OF DAILY TV VIEWING (LINEAR)
18 TVWATCHQ	0.97810	# HOURS OF DAILY TV VIEWING (QUADRATIC)
19 LANGHM23	0.80138	LANG OTHER THAN ENG AT HOME(SOME/ALWAYS)
20 HW-NO	0.97912	HOMEWORK (NONE ASSIGNED)
21 HW-YES	0.98018	HOMEWORK (YES, ASSIGNED)
22 HWLINEAR	0.91572	HOMEWORK (LINEAR)
23 HW:QUAD	0.87806	HOMEWORK (QUADRATIC)
24 PCTLUNCH	0.78587	PERCENT IN LUNCH PROGRAM (LINEAR)
25 %LUNCH M	0.86074	PERCENT IN LUNCH PROGRAM - MISSING
26 PCTWHT1	0.82509	PERCENT WHITE IN SCHOOL (0-49 %) WHT MIN
27 PCTWHT2	0.90899	PERCENT WHITE IN SCHOOL (50-79%) INTGRAT
28 AGE/GRD2	0.97629	AGE X GRADE: MODAL AGE, < MODAL GRADE
29 AGE/GRD3	0.98642	AGE X GRADE: MODAL AGE, MODAL GRADE,MISS
30 AGE/GRD4	0.00000	AGE X GRADE: MODAL AGE, > MODAL GRADE
31 AGE/GRD5	0.99087	AGE X GRADE: > MODAL AGE, MODAL GRADE
32 SCHTYPE	0.92757	SCHOOL TYPE: NON-PUBLIC
33 HW HELP	0.97190	SOMEONE AT HOME HELPS WITH HW:>ONCE WEEK
34 SINGLEP1	0.84599	HOW MANY PARENTS AT HOME? (BOTH)
35 MOMHOME1	0.79710	DOES YOUR MOTHER LIVE AT HOME? (YES)
36 MOMWORKY	0.88140	DOES YOUR MOTHER WORK FOR PAY? (YES)
37 PGS RD14	0.81218	PAGES/DAY READ FOR SCHOOL (6 OR MORE)
38 PGS RD13	0.80853	PAGES/DAY READ FOR SCHOOL (11 OR MORE)
39 SCHMSS12	0.95182	SCHOOL DAYS MISSED LAST MONTH (0-2 DAYS)
40 HS_PROG2	0.63909	HIGH SCHOOL PROGRAM (COLLEGE PREP)
41 HS_PROG3	0.90153	HIGH SCHOOL PROGRAM (VOC / TECH)
42 PHSPLAN2	0.88327	POST SECONDARY PLANS (2-YEAR COLLEGE)
43 PHSPLAN3	0.80474	POST SECONDARY PLANS (4-YEAR COLLEGE)
44 BEHAVRNM	0.84766	RULES OF BEHAVIOR ARE STRICT (NON-MISSG)
45 BEHAVR-L	0.89361	RULES OF BEHAVIOR ARE STRICT (LINEAR)
46 SAFETYNM	0.90438	DON'T FEEL SAFE AT SCHOOL (NON-MISSG)

Table F-19 (continued)
Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Mathematics Cross-sectional Conditioning Variables, Age 17/Grade 12

Contrast	Proportion of Variance	Description
47 SAFETY-L	0.62300	DON'T FEEL SAFE AT SCHOOL (LINEAR)
48 DISRPTNM	0.84616	STUDENTS OFTEN DISRUPT CLASS (NON-MISSG)
49 DISRPT-L	0.78183	STUDENTS OFTEN DISRUPT CLASS (LINEAR)
50 #SEM_ENG	0.86531	NUMBER OF SEMESTERS ENGLISH (NON MISSNG)
51 #SEMENGL	0.89239	NUMBER OF SEMESTERS OF ENGLISH (LINEAR)
52 #SEM_MAT	0.91089	NUMBER OF SEMESTERS OF MATH(NON MISSING)
53 #SEMMATL	0.90863	NUMBER OF SEMESTERS OF MATH (LINEAR)
54 #SEM_SCI	0.90242	NUMBER OF SEMESTERS SCIENCE(NON MISSING)
55 #SEMSCIL	0.85420	NUMBER OF SEMESTERS SCIENCE (LINEAR)
56 #SEM_HIS	0.91829	NUMBER OF SEMESTERS HIST (NON MISSING)
57 #SEMHISL	0.80940	NUMBER OF SEMESTERS OF HISTORY (LINEAR)
58 #SEMFLNG	0.92105	NUMBER OF SEMESTERS FOR.LANG(NON-MISSG)
59 #SEMFLGL	0.69453	NUMBER OF SEMESTERS FORGN LANG.(LINEAR)
60 #SEM_VOC	0.91069	NUMBER OF SEMESTERS VOC/TECH(NON-MISSG)
61 #SEMVOCL	0.89681	NUMBER OF SEMESTERS FORGN LANG.(LINEAR)
62 #SEM_ART	0.91626	NUMBER OF SEMESTERS ART/MUSIC(NON-MISS)
63 #SEMARTL	0.87657	NUMBER OF SEMESTERS ART/MUSIC (LINEAR)
64 SCHMATH	0.83474	SCHOOL LEVEL PROFICIENCY - MATH
65 SCHMATHM	0.00000	SCHOOL LEVEL PROFICIENCY MATH (MISSING)
66 SCHREAD	0.80642	SCHOOL LEVEL PROFICIENCY - READING
67 SCHREADM	0.99651	SCHOOL LEVEL PROFICIENCY READING (MISSG)
68 SCHSCI	0.83661	SCHOOL LEVEL PROFICIENCY - SCIENCE
69 SCHSCIM	0.99651	SCHOOL LEVEL PROFICIENCY SCIENCE (MISSG)
70 TEXTBOK1	0.92472	DO TEXTBOOK PROBLEMS (1-EVERY DAY)
71 TEXTBOK2	0.95984	DO TEXTBOOK PROBLEMS (2-SEVERAL/WEEK)
72 TEXTBOK3	0.98274	DO TEXTBOOK PROBLEMS (3-ONCE/WEEK)
73 TEXTBOK4	0.97228	DO TEXTBOOK PROBLEMS (4- <ONCE/WEEK)
74 WORKSHT1	0.95814	DO WORKSHEET PROBLEMS (1-EVERY DAY,MISS)
75 WORKSHT2	0.94645	DO WORKSHEET PROBLEMS (2-SEVERAL/WEEK)
76 WORKSHT3	0.94428	DO WORKSHEET PROBLEMS (3-ONCE/WEEK)
77 WORKSHT4	0.93863	DO WORKSHEET PROBLEMS (4-<ONCE/WEEK)
78 SMLLGRP1	0.91705	WORK IN SMALL GROUPS (1-EVERY DAY,MISS)
79 SMLLGRP2	0.92397	WORK IN SMALL GROUPS (2-SEVERAL/WEEK)
80 SMLLGRP3	0.90480	WORK IN SMALL GROUPS (3-ONCE/WEEK)
81 SMLLGRP4	0.89251	WORK IN SMALL GROUPS (4-<ONCE/WEEK)
82 W/OBJEC1	0.93120	WORK W/ SMALL OBJECTS(1-EVERY DAY,MISS)
83 W/OBJEC2	0.92906	WORK W/ SMALL OBJECTS(2-SEVERAL/WEEK)
84 W/OBJEC3	0.93940	WORK W/ SMALL OBJECTS(3-ONCE/WEEK)
85 W/OBJEC4	0.90104	WORK W/ SMALL OBJECTS(4-<ONCE/WEEK)
86 USECALC1	0.91959	USES CALCULATOR (1-EVERY DAY)
87 USECALC2	0.94778	USES CALCULATOR (2-SEVERAL/WEEK)
88 USECALC3	0.95321	USES CALCULATOR (3-ONCE/WEEK)
89 USECALC4	0.94367	USES CALCULATOR (4-<ONCE/WEEK)
90 USECOMP1	0.88179	USES COMPUTER (1-EVERY DAY,MISSING)
91 USECOMP2	0.92678	USES COMPUTER (2-SEVERAL/WEEK)

Table F-19 (continued)
Proportion of Variance of the Conditioning Variable Contrasts Accounted for
by the Principal Components Used in the Conditioning Model for
Mathematics Cross-sectional Conditioning Variables, Age 17/Grade 12

Contrast	Proportion of Variance	Description
92 USECOMP3	0.92982	USES COMPUTER (3-ONCE/WEEK)
93 USECOMP4	0.89252	USES COMPUTER (4-<ONCE/WEEK)
94 TAKTEST1	0.78120	TAKES TESTS (1-EVERY DAY,MISSING)
95 TAKTEST2	0.98025	TAKES TESTS (2-SEVERAL/WEEK)
96 TAKTEST3	0.95980	TAKES TESTS (3-ONCE/WEEK)
97 TAKTEST4	0.95472	TAKES TESTS (4-<ONCE/WEEK)
98 WRITREP1	0.72659	WRITES REPORTS (1-EVERY DAY,MISSING)
99 WRITREP2	0.94592	WRITES REPORTS (2-SEVERAL/WEEK)
100 WRITREP3	0.92768	WRITES REPORTS (3-ONCE/WEEK)
101 WRITREP4	0.88364	WRITES REPORTS (4-<ONCE/WEEK)
102 WRITPRF1	0.92085	WRITES PROOFS (1-EVERY DAY,MISSING)
103 WRITPRF2	0.93042	WRITES PROOFS (2-SEVERAL/WEEK)
104 WRITPRF3	0.92020	WRITES PROOFS (3-ONCE/WEEK)
105 WRITPRF4	0.89151	WRITES PROOFS (4-<ONCE/WEEK)
106 FORMPRB1	0.84806	FORMULATE PROBLEMS (1-EVERY DAY/MISSING)
107 FORMPRB2	0.92140	FORMULATE PROBLEMS (2-SEVERAL/WEEK)
108 FORMPRB3	0.91871	FORMULATE PROBLEMS (3-ONCE/WEEK)
109 FORMPRB4	0.89784	FORMULATE PROBLEMS (4-<ONCE/WEEK)
110 TCHEXPL1	0.95560	TEACHER EXPLAINS CALCULATOR USE (YES)
111 TCHEXPL2	0.95564	TEACHER EXPLAINS CALCULATOR USE (NO)
112 CALCCLS1	0.88977	USE CALCULATOR:CLASS PROBLMS(ALWAYS,MIS)
113 CALCCLS2	0.92801	USE CALCULATOR:CLASS PROBLEMS(SOMETIMES)
114 CALCHOM1	0.90830	USE CALCULATOR:HOME PROBLEMS(ALWAYS,MIS)
115 CALCHOM2	0.90834	USE CALCULATOR:HOME PROBLEMS (SOMETIMES)
116 CALCTST1	0.88155	USE CALCULATOR:ON TESTS (ALWAYS,MISSING)
117 CALCTST2	0.91504	USE CALCULATOR:ON TESTS (SOMETIMES)
118 MATHCLS1	0.95635	ARE YOU TAKING MATH CLASSES (YES)
119 MATHCLS2	0.94883	ARE YOU TAKING MATH CLASSES (NO)
120 HW DONE1	0.83404	AMOUNT OF HOMEWORK DONE (NONE)
121 HWDONE23	0.93505	AMOUNT OF HOMEWORK DONE (15-30 MINUTES)
122 HW DONE4	0.97107	AMOUNT OF HOMEWORK DONE (45 MINUTES)
123 HWDONE56	0.90687	AMOUNT OF HOMEWORK DONE (60, > HOUR)
124 GENRLMA1	0.97321	HOW LONG GENERAL MATH (> YEAR)
125 GENRLM23	0.98197	HOW LONG GENERAL MATH (1/2 - 1 YEAR)
126 GENRLMA4	0.97494	HOW LONG GENERAL MATH (NOT STUDIED)
127 CONSUMA1	0.98674	HOW LONG STUDIED CONSUMER MATH (> YEAR)
128 CONSUMA2	0.97603	HOW LONG STUDIED CONSUMER MATH (1 YEAR)
129 CONSUMA3	0.98691	HOW LONG STUDIED CONSUMER MATH(1/2 YEAR)
130 CONSUMA4	0.97534	HOW LONG STUD CONSUMER MATH(NOT STUDIED)
131 PREALGB1	0.95150	HOW LONG STUDIED PRE-ALGEBRA (> YEAR)
132 PREALGB2	0.97345	HOW LONG STUDIED PRE-ALGEBRA (1 YEAR)
133 PREALGB3	0.97217	HOW LONG STUDIED PRE-ALGEBRA (1/2 YEAR)
134 PREALGB4	0.97163	HOW LONG PRE-ALGEBRA (NOT STUDIED)
135 ALGEBRA1	0.83768	HOW LONG STUDIED ALGEBRA (> YEAR)
136 ALGEBRA2	0.96427	HOW LONG STUDIED ALGEBRA (1 YEAR)

Table F-19 (continued)
Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Mathematics Cross-sectional Conditioning Variables, Age 17/Grade 12

Contrast	Proportion of Variance	Description
137 ALGEBRA3	0.95407	HOW LONG STUDIED ALGEBRA (1/2 YEAR)
138 ALGEBRA4	0.88417	HOW LONG STUDIED ALGEBRA (NOT STUDIED)
139 GEOMRTY1	0.73255	HOW LONG STUDIED GEOMETRY (> YEAR)
140 GEOMRTY2	0.84403	HOW LONG STUDIED GEOMETRY (1 YEAR)
141 GEOMRTY3	0.98461	HOW LONG STUDIED GEOMETRY (1/2 YEAR)
142 GEOMRTY4	0.83213	HOW LONG STUDIED GEOMETRY (NOT STUDIED)
143 2NDALGB1	0.69643	HOW LONG STUDIED 2ND YR ALG(> YEAR)
144 2NDALGB2	0.87166	HOW LONG STUDIED 2ND YR ALG(1 YEAR)
145 2NDALGB3	0.91661	HOW LONG STUDIED 2ND YR ALG(1/2 YEAR)
146 2NDALGB4	0.82608	HOW LONG STUDIED 2ND YR ALG(NOT STUDIED)
147 TRIGMTY1	0.67788	HOW LONG STUDIED TRIGONOMETRY (> YEAR)
148 TRIGMTY2	0.94128	HOW LONG STUDIED TRIGONOMETRY (1 YEAR)
149 TRIGMTY3	0.82465	HOW LONG STUDIED TRIGONOMETRY (1/2 YEAR)
150 TRIGMTY4	0.84332	HOW LONG STUDIED TRIGONOMETRY (NOT STUD)
151 PRECALC1	0.75695	HOW LONG STUDIED PRE-CALCULUS (> YEAR)
152 PRECALC2	0.95248	HOW LONG STUDIED PRE-CALCULUS (1 YEAR)
153 PRECALC3	0.89441	HOW LONG STUDIED PRE-CALCULUS (1/2 YR)
154 PRECALC4	0.93492	HOW LONG STUDIED PRE-CALCULUS (NOT STUD)
155 PROBSTA1	0.94839	HOW LONG STUDIED PROB + STAT (> YEAR)
156 PROBSTA2	0.97829	HOW LONG STUDIED PROB + STAT (1 YEAR)
157 PROBSTA3	0.96766	HOW LONG STUDIED PROB + STAT (1/2 YEAR)
158 PROBSTA4	0.96323	HOW LONG STUDIED PROB + STAT (NOT STUD)
159 ANALGEO1	0.71424	HOW LONG STUDIED ANAL GEOMTRY (> YEAR)
160 ANALGEO2	0.97683	HOW LONG STUDIED ANAL GEOMTRY (1 YEAR)
161 ANALGEO3	0.92517	HOW LONG STUDIED ANAL GEOMTRY (1/2 YR)
162 ANALGEO4	0.94143	HOW LONG STUDIED ANAL GEOMTRY (NOT STUD)
163 CALCLUS1	0.82271	HOW LONG STUDIED CALCULUS (> YEAR)
164 CALCLUS2	0.93347	HOW LONG STUDIED CALCULUS (1 YEAR)
165 CALCLUS3	0.96749	HOW LONG STUDIED CALCULUS (1/2 YEAR)
166 CALCLUS4	0.95726	HOW LONG STUDIED CALCULUS (NOT STUDIED)
167 LIKMATH4	0.90466	LIKES MATHEMATICS (DISAGREE)
168 LIKMATH3	0.92472	LIKES MATHEMATICS (UNDECIDED)
169 LIKMATH2	0.94632	LIKES MATHEMATICS (AGREE)
170 LIKMATH1	0.85149	LIKES MATHEMATICS (STRONGLY AGREE)
171 USEMATH4	0.90914	USE MATHEMATICS (DISAGREE)
172 USEMATH3	0.93087	USE MATHEMATICS (UNDECIDED)
173 USEMATH2	0.97104	USE MATHEMATICS (AGREE)
174 USEMATH1	0.83090	USE MATHEMATICS (STRONGLY AGREE)
175 GOODMAT4	0.91604	GOOD AT MATHEMATICS (DISAGREE)
176 GOODMAT3	0.93723	GOOD AT MATHEMATICS (UNDECIDED)
177 GOODMAT2	0.95229	GOOD AT MATHEMATICS (AGREE)
178 GOODMAT1	0.83270	GOOD AT MATHEMATICS (STRONGLY AGREE)
179 MATBOY12	0.98217	MATHEMATICS IS FOR BOYS(AGREE/STR AGREE)
180 MATBOY34	0.96503	MATHEMATICS IS FOR BOYS(UNDECIDE/DISAGR)
181 MATBOYS5	0.95173	MATHEMATICS IS FOR BOYS(STRONGLY DISAGR)

Table F-19 (continued)
Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Mathematics Cross-sectional Conditioning Variables, Age 17/Grade 12

Contrast	Proportion of Variance	Description
182 USEFUL_4	0.91705	USEFUL FOR EVERYDAY PROBLEMS (DISAGREE)
183 USEFUL_3	0.93179	USEFUL FOR EVERYDAY PROBLEMS (UNDECIDED)
184 USEFUL_2	0.95598	USEFUL FOR EVERYDAY PROBLEMS (AGREE)
185 USEFUL_1	0.82714	USEFUL FOR EVERYDAY PROBLEMS (STR AGREE)
186 SEASON_W	0.98222	SEASON OF ASSESSMENT (WINTER)

Table F-20
Estimated Effects for the Mathematics Cross-sectional Principal Components
Age 17/Grade 12

Variable	Estimated Effect	Description
1 OVERALL	0.743221	OVERALL CONSTANT
2 PCFS1	0.033546	PRINCIPAL COMPONENT FACTOR SCORE 1
3 PCFS2	0.019398	PRINCIPAL COMPONENT FACTOR SCORE 2
4 PCFS3	-0.028442	PRINCIPAL COMPONENT FACTOR SCORE 3
5 PCFS4	-0.029522	PRINCIPAL COMPONENT FACTOR SCORE 4
6 PCFS5	-0.017012	PRINCIPAL COMPONENT FACTOR SCORE 5
7 PCFS6	0.010238	PRINCIPAL COMPONENT FACTOR SCORE 6
8 PCFS7	0.014310	PRINCIPAL COMPONENT FACTOR SCORE 7
9 PCFS8	-0.012487	PRINCIPAL COMPONENT FACTOR SCORE 8
10 PCFS9	-0.011828	PRINCIPAL COMPONENT FACTOR SCORE 9
11 PCFS10	-0.015515	PRINCIPAL COMPONENT FACTOR SCORE 10
12 PCFS11	-0.002829	PRINCIPAL COMPONENT FACTOR SCORE 11
13 PCFS12	-0.021406	PRINCIPAL COMPONENT FACTOR SCORE 12
14 PCFS13	-0.016244	PRINCIPAL COMPONENT FACTOR SCORE 13
15 PCFS14	0.030395	PRINCIPAL COMPONENT FACTOR SCORE 14
16 PCFS15	0.020173	PRINCIPAL COMPONENT FACTOR SCORE 15
17 PCFS16	-0.018073	PRINCIPAL COMPONENT FACTOR SCORE 16
18 PCFS17	-0.015395	PRINCIPAL COMPONENT FACTOR SCORE 17
19 PCFS18	-0.003483	PRINCIPAL COMPONENT FACTOR SCORE 18
20 PCFS19	0.003987	PRINCIPAL COMPONENT FACTOR SCORE 19
21 PCFS20	0.004597	PRINCIPAL COMPONENT FACTOR SCORE 20
22 PCFS21	-0.028230	PRINCIPAL COMPONENT FACTOR SCORE 21
23 PCFS22	0.018990	PRINCIPAL COMPONENT FACTOR SCORE 22
24 PCFS23	-0.008674	PRINCIPAL COMPONENT FACTOR SCORE 23
25 PCFS24	-0.019824	PRINCIPAL COMPONENT FACTOR SCORE 24
26 PCFS25	0.001339	PRINCIPAL COMPONENT FACTOR SCORE 25
27 PCFS26	0.030773	PRINCIPAL COMPONENT FACTOR SCORE 26
28 PCFS27	-0.007116	PRINCIPAL COMPONENT FACTOR SCORE 27
29 PCFS28	-0.003991	PRINCIPAL COMPONENT FACTOR SCORE 28
30 PCFS29	0.009891	PRINCIPAL COMPONENT FACTOR SCORE 29
31 PCFS30	-0.011847	PRINCIPAL COMPONENT FACTOR SCORE 30
32 PCFS31	-0.001194	PRINCIPAL COMPONENT FACTOR SCORE 31
33 PCFS32	-0.003706	PRINCIPAL COMPONENT FACTOR SCORE 32
34 PCFS33	-0.014489	PRINCIPAL COMPONENT FACTOR SCORE 33
35 PCFS34	0.000438	PRINCIPAL COMPONENT FACTOR SCORE 34
36 PCFS35	-0.014305	PRINCIPAL COMPONENT FACTOR SCORE 35
37 PCFS36	0.000899	PRINCIPAL COMPONENT FACTOR SCORE 36
38 PCFS37	0.008256	PRINCIPAL COMPONENT FACTOR SCORE 37
39 PCFS38	0.005121	PRINCIPAL COMPONENT FACTOR SCORE 38
40 PCFS39	-0.014618	PRINCIPAL COMPONENT FACTOR SCORE 39
41 PCFS40	0.019793	PRINCIPAL COMPONENT FACTOR SCORE 40
42 PCFS41	0.015551	PRINCIPAL COMPONENT FACTOR SCORE 41
43 PCFS42	-0.021057	PRINCIPAL COMPONENT FACTOR SCORE 42
44 PCFS43	-0.008893	PRINCIPAL COMPONENT FACTOR SCORE 43
45 PCFS44	0.000497	PRINCIPAL COMPONENT FACTOR SCORE 44
46 PCFS45	-0.000687	PRINCIPAL COMPONENT FACTOR SCORE 45
47 PCFS46	0.030423	PRINCIPAL COMPONENT FACTOR SCORE 46

Table F-20 (continued)
Estimated Effects for the Mathematics Cross-sectional Principal Components
Age 17/Grade 12

	Variable	Estimated Effect	Description
48	PCFS47	-0.009275	PRINCIPAL COMPONENT FACTOR SCORE 47
49	PCFS48	0.005454	PRINCIPAL COMPONENT FACTOR SCORE 48
50	PCFS49	-0.001820	PRINCIPAL COMPONENT FACTOR SCORE 49
51	PCFS50	-0.007607	PRINCIPAL COMPONENT FACTOR SCORE 50
52	PCFS51	-0.005916	PRINCIPAL COMPONENT FACTOR SCORE 51
53	PCFS52	-0.005836	PRINCIPAL COMPONENT FACTOR SCORE 52
54	PCFS53	0.013153	PRINCIPAL COMPONENT FACTOR SCORE 53
55	PCFS54	-0.007225	PRINCIPAL COMPONENT FACTOR SCORE 54
56	PCFS55	0.013115	PRINCIPAL COMPONENT FACTOR SCORE 55
57	PCFS56	0.015734	PRINCIPAL COMPONENT FACTOR SCORE 56
58	PCFS57	0.006198	PRINCIPAL COMPONENT FACTOR SCORE 57
59	PCFS58	-0.000034	PRINCIPAL COMPONENT FACTOR SCORE 58
60	PCFS59	0.003304	PRINCIPAL COMPONENT FACTOR SCORE 59
61	PCFS60	-0.006939	PRINCIPAL COMPONENT FACTOR SCORE 60
62	PCFS61	0.027668	PRINCIPAL COMPONENT FACTOR SCORE 61
63	PCFS62	-0.008318	PRINCIPAL COMPONENT FACTOR SCORE 62
64	PCFS63	0.019737	PRINCIPAL COMPONENT FACTOR SCORE 63
65	PCFS64	0.021824	PRINCIPAL COMPONENT FACTOR SCORE 64
66	PCFS65	0.026035	PRINCIPAL COMPONENT FACTOR SCORE 65
67	PCFS66	-0.018126	PRINCIPAL COMPONENT FACTOR SCORE 66
68	PCFS67	0.014186	PRINCIPAL COMPONENT FACTOR SCORE 67
69	PCFS68	0.012633	PRINCIPAL COMPONENT FACTOR SCORE 68
70	PCFS69	0.013955	PRINCIPAL COMPONENT FACTOR SCORE 69
71	PCFS70	-0.013062	PRINCIPAL COMPONENT FACTOR SCORE 70
72	PCFS71	-0.009145	PRINCIPAL COMPONENT FACTOR SCORE 71
73	PCFS72	0.006867	PRINCIPAL COMPONENT FACTOR SCORE 72
74	PCFS73	0.001924	PRINCIPAL COMPONENT FACTOR SCORE 73
75	PCFS74	0.006573	PRINCIPAL COMPONENT FACTOR SCORE 74
76	PCFS75	0.005545	PRINCIPAL COMPONENT FACTOR SCORE 75
77	PCFS76	-0.019417	PRINCIPAL COMPONENT FACTOR SCORE 76
78	PCFS77	0.010158	PRINCIPAL COMPONENT FACTOR SCORE 77
79	PCFS78	0.002177	PRINCIPAL COMPONENT FACTOR SCORE 78
80	PCFS79	-0.007116	PRINCIPAL COMPONENT FACTOR SCORE 79
81	PCFS80	-0.008674	PRINCIPAL COMPONENT FACTOR SCORE 80
82	PCFS81	-0.023141	PRINCIPAL COMPONENT FACTOR SCORE 81
83	PCFS82	-0.002388	PRINCIPAL COMPONENT FACTOR SCORE 82
84	PCFS83	0.029653	PRINCIPAL COMPONENT FACTOR SCORE 83
85	PCFS84	0.006920	PRINCIPAL COMPONENT FACTOR SCORE 84
86	PCFS85	0.000079	PRINCIPAL COMPONENT FACTOR SCORE 85
87	PCFS86	0.024179	PRINCIPAL COMPONENT FACTOR SCORE 86
88	PCFS87	0.006654	PRINCIPAL COMPONENT FACTOR SCORE 87
89	PCFS88	-0.013389	PRINCIPAL COMPONENT FACTOR SCORE 88
90	PCFS89	0.005577	PRINCIPAL COMPONENT FACTOR SCORE 89
91	PCFS90	-0.007151	PRINCIPAL COMPONENT FACTOR SCORE 90
92	PCFS91	-0.019238	PRINCIPAL COMPONENT FACTOR SCORE 91
93	PCFS92	0.004343	PRINCIPAL COMPONENT FACTOR SCORE 92
94	PCFS93	0.009632	PRINCIPAL COMPONENT FACTOR SCORE 93

Table F-20 (continued)
Estimated Effects for the Mathematics Cross-sectional Principal Components
Age 17/Grade 12

	Variable	Estimated Effect	Description
95	PCFS94	-0.019048	PRINCIPAL COMPONENT FACTOR SCORE 94
96	PCFS95	0.010954	PRINCIPAL COMPONENT FACTOR SCORE 95
97	PCFS96	0.000070	PRINCIPAL COMPONENT FACTOR SCORE 96
98	PCFS97	-0.004761	PRINCIPAL COMPONENT FACTOR SCORE 97
99	PCFS98	0.010021	PRINCIPAL COMPONENT FACTOR SCORE 98
100	PCFS99	-0.014143	PRINCIPAL COMPONENT FACTOR SCORE 99
101	PCFS100	0.018209	PRINCIPAL COMPONENT FACTOR SCORE 100
102	PCFS101	-0.000699	PRINCIPAL COMPONENT FACTOR SCORE 101
103	PCFS102	0.013365	PRINCIPAL COMPONENT FACTOR SCORE 102
104	PCFS103	0.005511	PRINCIPAL COMPONENT FACTOR SCORE 103
105	PCFS104	-0.014726	PRINCIPAL COMPONENT FACTOR SCORE 104
106	PCFS105	-0.023438	PRINCIPAL COMPONENT FACTOR SCORE 105
107	PCFS106	-0.003109	PRINCIPAL COMPONENT FACTOR SCORE 106
108	PCFS107	-0.000659	PRINCIPAL COMPONENT FACTOR SCORE 107
109	PCFS108	-0.003245	PRINCIPAL COMPONENT FACTOR SCORE 108
110	PCFS109	0.017181	PRINCIPAL COMPONENT FACTOR SCORE 109
111	PCFS110	0.024161	PRINCIPAL COMPONENT FACTOR SCORE 110

Table F-21
Estimated Effects for Mathematics Trend Conditioning Variable Contrasts, Age 9

Contrast	Estimated Effect	Description
1 OVERALL	-0.070947	OVERALL CONSTANT
2 GENDER2	-0.047264	SEX (FEMALE)
3 ETHNIC2	-0.149088	OBSERVED ETHNICITY (BLACK)
4 ETHNIC3	-0.310922	OBSERVED ETHNICITY (HISPANIC)
5 ETHNIC4	-0.520545	OBSERVED ETHNICITY (ASIAN)
6 STOC3	0.291426	SIZE AND TYPE OF COMMUNITY (HI METRO)
7 STOC1	0.131672	SIZE, TYPE COMMUN (NOT HI OR LO METRO)
8 REGION2	-0.067229	REGION (SOUTHEAST)
9 REGION3	-0.079322	REGION (CENTRAL)
10 REGION4	-0.089393	REGION (WEST)
11 PARED2	0.363150	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12 PARED3	0.521619	PARENTS EDUCATION (POST HIGH SCHOOL)
13 PARED4	0.563575	PARENTS EDUCATION (COLLEGE GRAD)
14 PARED_	0.288656	PARENTS EDUCATION (MISSING, I DON'T KNOW
15 MODGRD 1	-0.884012	< MODAL GRADE
16 MODGRD 2		> MODAL GRADE <<ZEROED OUT>>
17 HOMEITM2	0.191956	3 ITEMS IN THE HOME
18 HOMEITM3	0.351694	4 ITEMS IN THE HOME
19 RAC/SEX1	0.085632	BLACK, FEMALE
20 RAC/SEX2	-0.041830	HISPANIC, FEMALE
21 RAC/SEX3	0.306103	ASIAN AMERICAN, FEMALE
22 R/PE 1&2	-0.347279	BLACK, HS GRD & HS PLUS
23 NULL		BLACK, POST HIGH SCHOOL
24 RAC/PE 3	-0.460758	BLACK, COLLEGE GRAD
25 RAC/PE 4	-0.387415	BLACK, MISSING
26 R/PE 5&6	-0.120539	HISPANIC, HS GRD & HS PLUS
27 NULL		HISPANIC, POST HIGH SCHOOL
28 RAC/PE 7	-0.310826	HISPANIC, COLLEGE GRAD
29 RAC/PE 8	-0.115140	HISPANIC, MISSING
30 R/P10&11		AS AM HS GRD & HS PLUS <<ZEROED OUT>>
31 NULL		ASIAN AMERICAN, POST HIGH SCHOOL
32 RAC/PE11	0.116945	ASIAN AMERICAN, COLLEGE GRAD
33 RAC/PE12	-0.035154	ASIAN AMERICAN, MISSING
34 SCHTYPE	-0.001255	SCHOOL (NONPUBLIC)
35 LANGHM1	-0.279965	OTHER THAN ENGLISH AT HOME (SOMETIMES)
36 LANGHM2	-0.015385	OTHER THAN ENGLISH AT HOME (ALWAYS)
37 R/LN 1&2	-0.049266	BLACK, NOT ONLY ENGLISH AT HOME
38 NULL		BLACK, SOMETIMES OTHER THAN ENGLISH
39 R/LN 3&4	0.335220	HISP, NOT ONLY ENGLISH AT HOME
40 NULL		HISPANIC, SOMETIMES OTHER THAN ENGLISH
41 R/LN 5&6	0.278279	AS AM, NOT ONLY ENGLISH AT HOME
42 NULL		ASIAN AM, SOMETIMES OTHER THAN ENGLISH
43 DRACE2	-0.125166	DERIVED RACE (BLACK)
44 DRACE3	-0.308511	DERIVED RACE (HISPANIC)
45 DRACE4	0.125624	DERIVED RACE (ASIAN AMERICAN)

Table F-22
Estimated Effects for Mathematics Trend Conditioning Variable Contrasts, Age 13

Contrast	Estimated Effect	Description
1 OVERALL	-0.220824	OVERALL CONSTANT
2 GENDER2	-0.162231	SEX (FEMALE)
3 ETHNIC2	-0.249261	OBSERVED ETHNICITY (BLACK)
4 ETHNIC3	0.179249	OBSERVED ETHNICITY (HISPANIC)
5 ETHNIC4	-1.030550	OBSERVED ETHNICITY (ASIAN)
6 STOC3	0.353519	SIZE AND TYPE OF COMMUNITY (HIHG METRO)
7 STOC1	0.204794	SIZE & TYPE OF COMM (NOT HI, NOT LOW)
8 REGION2	-0.210375	REGION (SOUTHEAST)
9 REGION3	-0.077811	REGION (CENTRAL)
10 REGION4	-0.090462	REGION (WEST)
11 PARED2	0.078800	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12 PARED3	0.362890	PARENTS EDUCATION (POST HIGH SCHOOL)
13 PARED4	0.548685	PARENTS EDUCATION (COLLEGE GRAD)
14 PARED_	-0.105466	PARENTS EDUCATION (MISSING, I DON'T KNOW
15 MODLGRD1	-0.713239	< MODAL GRADE
16 MODLGRD2	0.302475	> MODAL GRADE
17 HOMEITM2	0.176835	3 ITEMS IN THE HOME
18 HOMEITM3	0.266237	4 ITEMS IN THE HOME
19 RAC/SEX1	0.204469	BLACK, FEMALE
20 RAC/SEX2	-0.053481	HISPANIC, FEMALE
21 RAC/SEX3	0.182785	ASIAN AMERICAN, FEMALE
22 RAC/PED1	0.020224	BLACK, HIGH SCHOOL GRAD
23 RAC/PED2	0.083533	BLACK, POST HIGH SCHOOL
24 RAC/PED3	-0.459795	BLACK, COLLEGE GRAD
25 RAC/PED4	0.027037	BLACK, MISSING
26 RAC/PED5	-0.181651	HISPANIC, HIGH SCHOOL GRAD
27 RAC/PED6	0.030523	HISPANIC, POST HIGH SCHOOL
28 RAC/PED7	-0.517543	HISPANIC, COLLEGE GRAD
29 RAC/PED8	-0.161294	HISPANIC, MISSING
30 RAC/PED9	0.166665	ASIAN AMERICAN, HIGH SCHOOL GRAD
31 RAC/PE10	0.648295	ASIAN AMERICAN, POST HIGH SCHOOL
32 RAC/PE11	0.575553	ASIAN AMERICAN, COLLEGE GRAD
33 RAC/PE12	0.446185	ASIAN AMERICAN, MISSING
34 SCHTYPE	-0.007142	SCHOOL (NONPUBLIC)
35 HW-NO	-0.016534	HOMEWORK (NONE ASSIGNED)
36 HW-YES	0.028207	HOMEWORK (YES - DIDN'T DO)
37 HW-234	0.025499	HOMEWORK (1/2 HR TO 2 HOURS)
38 LANGHM1	0.169194	OTHER THAN ENGLISH AT HOME (SOMETIMES)
39 LANGHM2	-0.100079	OTHER THAN ENGLISH AT HOME (ALWAYS)
40 RAC/LNG1	0.054392	BLACK, ALWAYS OTHER THAN ENGLISH
41 RAC/LNG2	0.021513	BLACK, SOMETIMES OTHER THAN ENGLISH
42 RAC/LNG3	0.176333	HISPANIC, ALWAYS OTHER THAN ENGLISH
43 RAC/LNG4	0.064036	HISPANIC, SOMETIMES OTHER THAN ENGLISH
44 RAC/LNG5	0.556046	ASIAN AM, ALWAYS OTHER THAN ENGLISH
45 RAC/LNG6	0.085855	ASIAN AM, SOMETIMES OTHER THAN ENGLISH
46 DRACE2	-0.375643	DERIVED RACE (BLACK)
47 DRACE3	-0.472708	DERIVED RACE (HISPANIC)
48 DRACE4	0.303911	DERIVED RACE (ASIAN AMERICAN)

Table F-23
Estimated Effects for Mathematics Trend Conditioning Variable Contrasts, Age 17

Contrast	Estimated Effect	Description
1 OVERALL	-0.993268	OVERALL CONSTANT
2 GENDER2	-0.246866	SEX (FEMALE)
3 ETHNIC2	-0.319910	OBSERVED ETHNICITY (BLACK)
4 ETHNIC3	-0.073881	OBSERVED ETHNICITY (HISPANIC)
5 ETHNIC4	-0.343919	OBSERVED ETHNICITY (ASIAN)
6 STOC3	0.326225	SIZE AND TYPE OF COMMUNITY (HIHG METRO)
7 STOC1	0.198616	SIZE & TYPE OF COMM (NOT HI, NOT LOW)
8 REGION2	0.073790	REGION (SOUTHEAST)
9 REGION3	0.178219	REGION (CENTRAL)
10 REGION4	0.053721	REGION (WEST)
11 PARED2	0.080868	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12 PARED3	0.224718	PARENTS EDUCATION (POST HIGH SCHOOL)
13 PARED4	0.297060	PARENTS EDUCATION (COLLEGE GRAD)
14 PARED_	-0.198044	PARENTS EDUCATION (MISSING, I DON'T KNOW
15 MODLGRD1	-0.270404	< MODAL GRADE
16 MODLGRD2	-0.031149	> MODAL GRADE
17 HOMEITM2	0.009241	3 ITEMS IN THE HOME
18 HOMEITM3	0.051571	4 ITEMS IN THE HOME
19 RAC/SEX1	0.104469	BLACK, FEMALE
20 RAC/SEX2	0.180426	HISPANIC, FEMALE
21 RAC/SEX3	0.223051	ASIAN AMERICAN, FEMALE
22 RAC/PED1	-0.025858	BLACK, HIGH SCHOOL GRAD
23 RAC/PED2	-0.053037	BLACK, POST HIGH SCHOOL
24 RAC/PED3	-0.077071	BLACK, COLLEGE GRAD
25 RAC/PED4	0.291763	BLACK, MISSING
26 RAC/PED5	-0.252616	HISPANIC, HIGH SCHOOL GRAD
27 RAC/PED6	0.106646	HISPANIC, POST HIGH SCHOOL
28 RAC/PED7	-0.284662	HISPANIC, COLLEGE GRAD
29 RAC/PED8	-0.408317	HISPANIC, MISSING
30 RAC/PED9		ASIAN AMERICAN, HIGH SCH GRAD (ZEROED OU
31 RAC/PE10	-0.313759	ASIAN AMERICAN, POST HIGH SCH OR COL GRD
32 RAC/PE11		ASIAN AMERICAN, COL GRD (MERGED WITH ABO
33 RAC/PE12		ASIAN AMERICAN, MISSING (ZEROED OUT)
34 SCHTYPE	0.016251	SCHOOL (NONPUBLIC)
35 HW-NO	-0.314791	HOMEWORK (NONE ASSIGNED)
36 HW-YES	-0.116352	HOMEWORK (YES)
37 HW-234	-0.003032	HOMEWORK (1/2 HR TO 2 HOURS)
38 LANGHM1	-0.027535	OTHER THAN ENGLISH AT HOME (SOMETIMES)
39 LANGHM2	-0.117091	OTHER THAN ENGLISH AT HOME (ALWAYS)
40 RAC/LNG1	0.131059	BLACK, ALWAYS OTHER THAN ENGLISH
41 RAC/LNG2	0.068216	BLACK, SOMETIMES OTHER THAN ENGLISH
42 RAC/LNG3	-0.166607	HISPANIC, ALWAYS OTHER THAN ENGLISH
43 RAC/LNG4	-0.164103	HISPANIC, SOMETIMES OTHER THAN ENGLISH
44 RAC/LNG5	0.467924	ASIAN AM, ALWAYS OTHER THAN ENGLISH
45 RAC/LNG6	0.806636	ASIAN AM, SOMETIMES OTHER THAN ENGLISH
46 NMATH1	0.088480	PRE-ALGEBRA
47 NMATH2	0.446946	ALGEBRA
48 NMATH3	0.649685	GEOMETRY
49 NMATH4	1.178459	ALGEBRA 2

Table F-23 (continued)
Estimated Effects for Mathematics Trend Conditioning Variable Contrasts, Age 17

Contrast	Estimated Effect	Description
50 NMATH5	1.852285	CALCULUS
51 HS_PGM2	0.233761	COLL PREP
52 HS_PGM3	-0.033956	VOCAT/TECH
53 DRACE2	-0.197809	DERIVED RACE (BLACK)
54 DRACE3	-0.085765	DERIVED RACE (HISPANIC)
55 DRACE4	0.027766	DERIVED RACE (ASIAN AMERICAN)

Table F-24
Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Science Cross-sectional Conditioning Variables, Age 9/Grade 4

Contrast	Proportion of Variance	Description
2 GENDER2	0.86656	SEX (FEMALE)
3 ETHNIC2	0.68266	DERIVED RACE (BLACK)
4 ETHNIC3	0.77871	DERIVED RACE (HISPANIC)
5 ETHNIC4	0.92591	DERIVED RACE (ASIAN)
6 STOC3	0.86974	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7 STOC147	0.88555	SIZE AND TYPE OF COMMUNITY (NON-HI+LOW)
8 REGION2	0.90937	REGION (SOUTHEAST)
9 REGION3	0.90106	REGION (CENTRAL)
10 REGION4	0.89360	REGION (WEST)
11 PARED2	0.97749	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12 PARED3	0.98723	PARENTS EDUCATION (POST HIGH SCHOOL)
13 PARED4	0.96328	PARENTS EDUCATION (COLLEGE GRAD)
14 PARED_	0.96257	PARENTS EDUCATION (MISSING,I DON'T KNOW)
15 HOMEITM2	0.81286	3 ITEMS IN THE HOME
16 HOMEITM3	0.76264	4 ITEMS IN THE HOME
17 TVWATCHL	0.98722	# HOURS OF DAILY TV VIEWING (LINEAR)
18 TVWATCHQ	0.98757	# HOURS OF DAILY TV VIEWING (QUADRATIC)
19 LANGHM23	0.75788	LANG OTHER THAN ENG AT HOME(SOME/ALWAYS)
20 HW-NO	0.86623	HOMEWORK (NONE ASSIGNED)
21 HW-YES	0.87514	HOMEWORK (YES, ASSIGNED)
22 HWLINEAR	0.92272	HOMEWORK (LINEAR)
23 HW:QUAD	0.81786	HOMEWORK (QUADRATIC)
24 PCTLUNCH	0.79562	PERCENT IN LUNCH PROGRAM (LINEAR)
25 %LUNCH M	0.88796	PERCENT IN LUNCH PROGRAM - MISSING
26 PCTWHT1	0.78661	PERCENT WHITE IN SCHOOL (0-49 %) WHT MIN
27 PCTWHT2	0.91936	PERCENT WHITE IN SCHOOL (50-79%) INTGRAT
28 AGE/GRD2	0.95938	AGE X GRADE: MODAL AGE, < MODAL GRADE
29 AGE/GRD3	0.98283	AGE X GRADE: MODAL AGE, MODAL GRADE,MISS
30 AGE/GRD4	0.99063	AGE X GRADE: MODAL AGE, > MODAL GRADE
31 AGE/GRD5	0.98892	AGE X GRADE: > MODAL AGE, MODAL GRADE
32 SCHTYPE	0.87549	SCHOOL TYPE: NON-PUBLIC
33 HW HELP	0.88900	SOMEONE AT HOME HELPS WITH HW:>ONCE WEEK
34 PRESCH_Y	0.87414	WENT TO PRESCHOOL? (YES)
35 SINGLEP1	0.80352	HOW MANY PARENTS AT HOME? (BOTH)
36 MOMHOME1	0.79806	DOES YOUR MOTHER LIVE AT HOME? (YES)
37 MOMWORKY	0.94200	DOES YOUR MOTHER WORK FOR PAY? (YES)
38 PGS RD14	0.82607	PAGES/DAY READ FOR SCHOOL (6 OR MORE)
39 PGS RD13	0.82894	PAGES/DAY READ FOR SCHOOL (11 OR MORE)
40 SCHMATH	0.74698	SCHOOL LEVEL PROFICIENCY - MATH
41 SCHMATHM	0.70695	SCHOOL LEVEL PROFICIENCY MATH (MISSING)
42 SCHREAD	0.74603	SCHOOL LEVEL PROFICIENCY - READING
43 SCHREADM	0.91667	SCHOOL LEVEL PROFICIENCY READING (MISSG)
44 SCHSCI	0.77135	SCHOOL LEVEL PROFICIENCY - SCIENCE
45 SCHSCIM	0.92640	SCHOOL LEVEL PROFICIENCY SCIENCE (MISSG)
46 P/A_EXP1	0.98833	PLANT/ANIMAL EXPERIMENTS (YES)

Table F-24 (continued)
Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Science Cross-sectional Conditioning Variables, Age 9/Grade 4

Contrast	Proportion of Variance	Description
47 P/A_EXP2	0.98818	PLANT/ANIMAL EXPERIMENTS (NO)
48 ELE_EXP1	0.98325	ELECTRICAL EXPERIMENTS (YES)
49 ELE_EXP2	0.98398	ELECTRICAL EXPERIMENTS (NO)
50 CHM_EXP1	0.98004	CHEMICAL EXPERIMENTS (YES)
51 CHM_EXP2	0.98104	CHEMICAL EXPERIMENTS (NO)
52 R/M_EXP1	0.98220	ROCK/MINERAL EXPERIMENTS (YES)
53 R/M_EXP2	0.98313	ROCK/MINERAL EXPERIMENTS (NO)
54 TEL_EXP1	0.98009	TELESCOPE EXPERIMENTS (YES)
55 TEL_EXP2	0.98167	TELESCOPE EXPERIMENTS (NO)
56 WEA_EXP1	0.97593	WEATHER EXPERIMENTS (YES)
57 WEA_EXP2	0.97995	WEATHER EXPERIMENTS (NO)
58 LIKESCI1	0.97259	LIKE SCIENCE? (YES)
59 LIKESCI2	0.96185	LIKE SCIENCE? (NO)
60 AMTSCI1	0.89042	AMOUNT OF SCIENCE (1-EVERY DA)
61 AMTSCI2	0.94576	AMOUNT OF SCIENCE (2-SEVERAL/WEEK)
62 AMTSCI3	0.82641	AMOUNT OF SCIENCE (3-ONCE/WEEK)
63 AMTSCI4	0.91018	AMOUNT OF SCIENCE (4-< ONCE/WEEK)
64 AMTSCI5	0.71539	AMOUNT OF SCIENCE (5-NEVER)
65 HW_DONE1	0.97861	AMOUNT OF HOMEWORK DONE (1-NONE)
66 HW_DONE2	0.95317	AMOUNT OF HOMEWORK DONE (2-1/2 HOUR)
67 HW_DONE3	0.96775	AMOUNT OF HOMEWORK DONE (3-ONE HOUR)
68 HW_DONE4	0.95996	AMOUNT OF HOMEWORK DONE (4-TWO HOURS)
69 HW_DONE5	0.94526	AMOUNT OF HOMEWORK DONE (5-> 2 HOURS)
70 HW_DOME6	0.75448	AMOUNT OF HOMEWORK DONE (6-NO SCI CLASS)
71 WEEKLPJ1	0.95536	WEEK-LONG PROJECTS (YES)
72 WEEKLPJ2	0.96390	WEEK-LONG PROJECTS (NO)
73 READTXT1	0.83242	READ A TEXTBOOK (1-EVERYDAY)
74 READTXT2	0.94764	READ A TEXTBOOK (2-SEVERAL/WEEK)
75 READTXT3	0.86062	READ A TEXTBOOK (3-ONCE/WEEK)
76 READTXT4	0.90092	READ A TEXTBOOK (4-< ONCE/WEEK)
77 READTXT5	0.91802	READ A TEXTBOOK (5-NEVER)
78 SCINEWS1	0.96238	DISCUSS SCIENCE NEWS (1-EVERYDAY)
79 SCINEWS2	0.95960	DISCUSS SCIENCE NEWS (2-SEVERAL/WEEK)
80 SCINEWS3	0.97411	DISCUSS SCIENCE NEWS (3-ONCE/WEEK)
81 SCINEWS4	0.96650	DISCUSS SCIENCE NEWS (4-< ONCE/WEEK)
82 SCINEWS5	0.96021	DISCUSS SCIENCE NEWS (5-NEVER)
83 WORKOTH1	0.77957	WORK WITH OTHERS (1-EVERYDAY)
84 WORKOTH2	0.94584	WORK WITH OTHERS (2-SEVERAL/WEEK)
85 WORKOTH3	0.97327	WORK WITH OTHERS (3-ONCE/WEEK)
86 WORKOTH4	0.93114	WORK WITH OTHERS (4-< ONCE/WEEK)
87 WORKOTH5	0.88791	WORK WITH OTHERS (5-NEVER)
88 REPORTS1	0.80124	GIVE REPORTS (1-EVERYDAY)
89 REPORTS2	0.91533	GIVE REPORTS (2-SEVERAL/WEEK)
90 REPORTS3	0.95668	GIVE REPORTS (3-ONCE/WEEK)
91 REPORTS4	0.88898	GIVE REPORTS (4-< ONCE/WEEK)

Table F-24 (continued)
Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Science Cross-sectional Conditioning Variables, Age 9/Grade 4

Contrast	Proportion of Variance	Description
92 REPORTS5	0.91088	GIVE REPORTS (5-NEVER)
93 EXPERMT1	0.74415	DO EXPERIMENTS (1-EVERYDAY)
94 EXPERMT2	0.95576	DO EXPERIMENTS (2-SEVERAL/WEEK)
95 EXPERMT3	0.96756	DO EXPERIMENTS (3-ONCE/WEEK)
96 EXPERMT4	0.91806	DO EXPERIMENTS (4-< ONCE/WEEK)
97 EXPERMT5	0.88891	DO EXPERIMENTS (5-NEVER)
98 SEASON_W	0.93618	SEASON - WINTER

Table F-25
Estimated Effects for the Science Cross-sectional Principal Components
Age 9/Grade 4

Variable	Estimated Effect	Description
1 OVERALL	-0.893197	OVERALL CONSTANT
2 PCFS1	0.074837	PRINCIPAL COMPONENT FACTOR SCORE 1
3 PCFS2	-0.055639	PRINCIPAL COMPONENT FACTOR SCORE 2
4 PCFS3	0.042624	PRINCIPAL COMPONENT FACTOR SCORE 3
5 PCFS4	0.053510	PRINCIPAL COMPONENT FACTOR SCORE 4
6 PCFS5	0.005152	PRINCIPAL COMPONENT FACTOR SCORE 5
7 PCFS6	0.014997	PRINCIPAL COMPONENT FACTOR SCORE 6
8 PCFS7	0.020088	PRINCIPAL COMPONENT FACTOR SCORE 7
9 PCFS8	0.044597	PRINCIPAL COMPONENT FACTOR SCORE 8
10 PCFS9	0.011401	PRINCIPAL COMPONENT FACTOR SCORE 9
11 PCFS10	0.001260	PRINCIPAL COMPONENT FACTOR SCORE 10
12 PCFS11	0.041303	PRINCIPAL COMPONENT FACTOR SCORE 11
13 PCFS12	-0.040567	PRINCIPAL COMPONENT FACTOR SCORE 12
14 PCFS13	0.012151	PRINCIPAL COMPONENT FACTOR SCORE 13
15 PCFS14	0.009763	PRINCIPAL COMPONENT FACTOR SCORE 14
16 PCFS15	-0.007513	PRINCIPAL COMPONENT FACTOR SCORE 15
17 PCFS16	-0.006695	PRINCIPAL COMPONENT FACTOR SCORE 16
18 PCFS17	-0.018072	PRINCIPAL COMPONENT FACTOR SCORE 17
19 PCFS18	-0.011101	PRINCIPAL COMPONENT FACTOR SCORE 18
20 PCFS19	-0.009648	PRINCIPAL COMPONENT FACTOR SCORE 19
21 PCFS20	-0.020165	PRINCIPAL COMPONENT FACTOR SCORE 20
22 PCFS21	0.019111	PRINCIPAL COMPONENT FACTOR SCORE 21
23 PCFS22	-0.006203	PRINCIPAL COMPONENT FACTOR SCORE 22
24 PCFS23	0.019017	PRINCIPAL COMPONENT FACTOR SCORE 23
25 PCFS24	-0.006786	PRINCIPAL COMPONENT FACTOR SCORE 24
26 PCFS25	0.027186	PRINCIPAL COMPONENT FACTOR SCORE 25
27 PCFS26	0.025310	PRINCIPAL COMPONENT FACTOR SCORE 26
28 PCFS27	-0.023397	PRINCIPAL COMPONENT FACTOR SCORE 27
29 PCFS28	-0.001404	PRINCIPAL COMPONENT FACTOR SCORE 28
30 PCFS29	0.011209	PRINCIPAL COMPONENT FACTOR SCORE 29
31 PCFS30	0.008481	PRINCIPAL COMPONENT FACTOR SCORE 30
32 PCFS31	0.012216	PRINCIPAL COMPONENT FACTOR SCORE 31
33 PCFS32	-0.004200	PRINCIPAL COMPONENT FACTOR SCORE 32
34 PCFS33	0.012409	PRINCIPAL COMPONENT FACTOR SCORE 33
35 PCFS34	0.043750	PRINCIPAL COMPONENT FACTOR SCORE 34
36 PCFS35	0.003100	PRINCIPAL COMPONENT FACTOR SCORE 35
37 PCFS36	0.005367	PRINCIPAL COMPONENT FACTOR SCORE 36
38 PCFS37	-0.016649	PRINCIPAL COMPONENT FACTOR SCORE 37
39 PCFS38	-0.022812	PRINCIPAL COMPONENT FACTOR SCORE 38
40 PCFS39	0.022547	PRINCIPAL COMPONENT FACTOR SCORE 39
41 PCFS40	0.005027	PRINCIPAL COMPONENT FACTOR SCORE 40
42 PCFS41	0.015879	PRINCIPAL COMPONENT FACTOR SCORE 41
43 PCFS42	0.017001	PRINCIPAL COMPONENT FACTOR SCORE 42
44 PCFS43	-0.004826	PRINCIPAL COMPONENT FACTOR SCORE 43
45 PCFS44	-0.025154	PRINCIPAL COMPONENT FACTOR SCORE 44
46 PCFS45	-0.038651	PRINCIPAL COMPONENT FACTOR SCORE 45
47 PCFS46	0.011810	PRINCIPAL COMPONENT FACTOR SCORE 46

Table F-25 (continued)
Estimated Effects for the Science Cross-sectional Principal Components
Age 9/Grade 4

	Variable	Estimated Effect	Description
48	PCFS47	-0.020072	PRINCIPAL COMPONENT FACTOR SCORE 47
49	PCFS48	0.002069	PRINCIPAL COMPONENT FACTOR SCORE 48
50	PCFS49	-0.012889	PRINCIPAL COMPONENT FACTOR SCORE 49
51	PCFS50	-0.018576	PRINCIPAL COMPONENT FACTOR SCORE 50
52	PCFS51	0.002290	PRINCIPAL COMPONENT FACTOR SCORE 51
53	PCFS52	0.005040	PRINCIPAL COMPONENT FACTOR SCORE 52
54	PCFS53	0.042905	PRINCIPAL COMPONENT FACTOR SCORE 53
55	PCFS54	-0.024997	PRINCIPAL COMPONENT FACTOR SCORE 54
56	PCFS55	0.036833	PRINCIPAL COMPONENT FACTOR SCORE 55
57	PCFS56	0.013612	PRINCIPAL COMPONENT FACTOR SCORE 56
58	PCFS57	-0.006963	PRINCIPAL COMPONENT FACTOR SCORE 57
59	PCFS58	0.046667	PRINCIPAL COMPONENT FACTOR SCORE 58
60	PCFS59	0.024425	PRINCIPAL COMPONENT FACTOR SCORE 59

Table F-26
Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Science Cross-sectional Conditioning Variables, Age 13/Grade 8

Contrast	Proportion of Variance	Description
2 GENDER2	0.88317	SEX (FEMALE)
3 ETHNIC2	0.86074	DERIVED RACE (BLACK)
4 ETHNIC3	0.80408	DERIVED RACE (HISPANIC)
5 ETHNIC4	0.87475	DERIVED RACE (ASIAN)
6 STOC3	0.87382	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7 STOC147	0.88913	SIZE AND TYPE OF COMMUNITY (NON-HI+LOW)
8 REGION2	0.88872	REGION (SOUTHEAST)
9 REGION3	0.88473	REGION (CENTRAL)
10 REGION4	0.87188	REGION (WEST)
11 PARED2	0.89787	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12 PARED3	0.89490	PARENTS EDUCATION (POST HIGH SCHOOL)
13 PARED4	0.88637	PARENTS EDUCATION (COLLEGE GRAD)
14 PARED_	0.90938	PARENTS EDUCATION (MISSING,I DON'T KNOW)
15 HOMEITM2	0.86301	3 ITEMS IN THE HOME
16 HOMEITM3	0.83580	4 ITEMS IN THE HOME
17 TVWATCHL	0.98409	# HOURS OF DAILY TV VIEWING (LINEAR)
18 TVWATCHQ	0.98411	# HOURS OF DAILY TV VIEWING (QUADRATIC)
19 LANGHM23	0.77285	LANG OTHER THAN ENG AT HOME(SOME/ALWAYS)
20 HW-NO	0.97059	HOMEWORK (NONE ASSIGNED)
21 HW-YES	0.97140	HOMEWORK (YES, ASSIGNED)
22 HWLINEAR	0.96797	HOMEWORK (LINEAR)
23 HW:QUAD	0.96544	HOMEWORK (QUADRATIC)
24 PCTLUNCH	0.83171	PERCENT IN LUNCH PROGRAM (LINEAR)
25 %LUNCH M	0.82495	PERCENT IN LUNCH PROGRAM - MISSING
26 PCTWHT1	0.80071	PERCENT WHITE IN SCHOOL (0-49 %) WHT MIN
27 PCTWHT2	0.83456	PERCENT WHITE IN SCHOOL (50-79%) INTGRAT
28 AGE/GRD2	0.79423	AGE X GRADE: MODAL AGE, < MODAL GRADE
29 AGE/GRD3	0.93773	AGE X GRADE: MODAL AGE, MODAL GRADE,MISS
30 AGE/GRD4	0.80349	AGE X GRADE: MODAL AGE, > MODAL GRADE
31 AGE/GRD5	0.94437	AGE X GRADE: > MODAL AGE, MODAL GRADE
32 SCHTYPE	0.83047	SCHOOL TYPE: NON-PUBLIC
33 HW HELP	0.96902	SOMEONE AT HOME HELPS WITH HW:>ONCE WEEK
34 SINGLEP1	0.72362	HOW MANY PARENTS AT HOME? (BOTH)
35 MOMHOME1	0.80172	DOES YOUR MOTHER LIVE AT HOME? (YES)
36 MOMWORKY	0.63532	DOES YOUR MOTHER WORK FOR PAY? (YES)
37 PGS RD14	0.79605	PAGES/DAY READ FOR SCHOOL (6 OR MORE)
38 PGS RD13	0.79868	PAGES/DAY READ FOR SCHOOL (11 OR MORE)
39 EXPHSGRD	0.65991	EXPECT TO GRADUATE FROM HIGH SCHOOL(YES)
40 SCHMSS12	0.95985	SCHOOL DAYS MISSED LAST MONTH (0-2 DAYS)
41 BEHAVRNM	0.81193	RULES OF BEHAVIOR ARE STRICT (NON-MISSG)
42 BEHAVR-L	0.84965	RULES OF BEHAVIOR ARE STRICT (LINEAR)
43 SAFETYNM	0.86437	DON'T FEEL SAFE AT SCHOOL (NON-MISSG)
44 SAFETY-L	0.67017	DON'T FEEL SAFE AT SCHOOL (LINEAR)
45 DISRPTNM	0.85724	STUDENTS OFTEN DISRUPT CLASS (NON-MISSG)
46 DISRPT-L	0.71047	STUDENTS OFTEN DISRUPT CLASS (LINEAR)

Table F-26 (continued)
Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Science Cross-sectional Conditioning Variables, Age 13/Grade 8

Contrast	Proportion of Variance	Description
47 SCHMATH	0.79217	SCHOOL LEVEL PROFICIENCY - MATH
48 SCHMATHM	0.76390	SCHOOL LEVEL PROFICIENCY MATH (MISSING)
49 SCHREAD	0.81946	SCHOOL LEVEL PROFICIENCY - READING
50 SCHREADM	0.95184	SCHOOL LEVEL PROFICIENCY READING (MISSG)
51 SCHSCI	0.81350	SCHOOL LEVEL PROFICIENCY - SCIENCE
52 SCHSCIM	0.94935	SCHOOL LEVEL PROFICIENCY SCIENCE (MISSG)
53 P/A_EXP1	0.99528	PLANT/ANIMAL EXPERIMENTS (YES)
54 P/A_EXP2	0.99542	PLANT/ANIMAL EXPERIMENTS (NO)
55 ELE_EXP1	0.99577	ELECTRICAL EXPERIMENTS (YES)
56 ELE_EXP2	0.99540	ELECTRICAL EXPERIMENTS (NO)
57 CHM_EXP1	0.99288	CHEMICAL EXPERIMENTS (YES)
58 CHM_EXP2	0.99291	CHEMICAL EXPERIMENTS (NO)
59 R/M_EXP1	0.99353	ROCK/MINERAL EXPERIMENTS (YES)
60 R/M_EXP2	0.99365	ROCK/MINERAL EXPERIMENTS (NO)
61 TEL_EXP1	0.99460	TELESCOPE EXPERIMENTS (YES)
62 TEL_EXP2	0.99476	TELESCOPE EXPERIMENTS (NO)
63 WEA_EXP1	0.99441	WEATHER EXPERIMENTS (YES)
64 WEA_EXP2	0.99445	WEATHER EXPERIMENTS (NO)
65 LIKESCI1	0.99349	LIKE SCIENCE? (YES)
66 LIKESCI2	0.99324	LIKE SCIENCE? (NO)
67 AMTSCI1	0.97657	AMOUNT OF SCIENCE (1-EVERY DA)
68 AMTSCI2	0.97630	AMOUNT OF SCIENCE (2-SEVERAL/WEEK)
69 AMTSCI3	0.97876	AMOUNT OF SCIENCE (3-ONCE/WEEK)
70 AMTSCI4	0.98344	AMOUNT OF SCIENCE (4-< ONCE/WEEK)
71 AMTSCI5	0.81507	AMOUNT OF SCIENCE (5-NEVER)
72 HW_DONE1	0.93814	AMOUNT OF HOMEWORK DONE (1-NONE)
73 HW_DONE2	0.98838	AMOUNT OF HOMEWORK DONE (2-1/2 HOUR)
74 HW_DONE3	0.98426	AMOUNT OF HOMEWORK DONE (3-ONE HOUR)
75 HW_DONE4	0.97743	AMOUNT OF HOMEWORK DONE (4-TWO HOURS)
76 HW_DONE5	0.94869	AMOUNT OF HOMEWORK DONE (5-> 2 HOURS)
77 HW_DOME6	0.80300	AMOUNT OF HOMEWORK DONE (6-NO SCI CLASS)
78 WEEKLPJ1	0.99143	WEEK-LONG PROJECTS (YES)
79 WEEKLPJ2	0.99131	WEEK-LONG PROJECTS (NO)
80 READTXT1	0.95387	READ A TEXTBOOK (1-EVERYDAY)
81 READTXT2	0.97660	READ A TEXTBOOK (2-SEVERAL/WEEK)
82 READTXT3	0.99050	READ A TEXTBOOK (3-ONCE/WEEK)
83 READTXT4	0.98164	READ A TEXTBOOK (4-< ONCE/WEEK)
84 READTXT5	0.72461	READ A TEXTBOOK (5-NEVER)
85 SCINEWS1	0.95143	DISCUSS SCIENCE NEWS (1-EVERYDAY)
86 SCINEWS2	0.96185	DISCUSS SCIENCE NEWS (2-SEVERAL/WEEK)
87 SCINEWS3	0.98946	DISCUSS SCIENCE NEWS (3-ONCE/WEEK)
88 SCINEWS4	0.98467	DISCUSS SCIENCE NEWS (4-< ONCE/WEEK)
89 SCINEWS5	0.87052	DISCUSS SCIENCE NEWS (5-NEVER)
90 WORKOTH1	0.90571	WORK WITH OTHERS (1-EVERYDAY)
91 WORKOTH2	0.97765	WORK WITH OTHERS (2-SEVERAL/WEEK)

Table F-26 (continued)
Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Science Cross-sectional Conditioning Variables, Age 13/Grade 8

Contrast	Proportion of Variance	Description
92 WORKOTH3	0.98939	WORK WITH OTHERS (3-ONCE/WEEK)
93 WORKOTH4	0.98433	WORK WITH OTHERS (4-< ONCE/WEEK)
94 WORKOTH5	0.86163	WORK WITH OTHERS (5-NEVER)
95 REPORTS1	0.93020	GIVE REPORTS (1-EVERYDAY)
96 REPORTS2	0.96502	GIVE REPORTS (2-SEVERAL/WEEK)
97 REPORTS3	0.97802	GIVE REPORTS (3-ONCE/WEEK)
98 REPORTS4	0.98150	GIVE REPORTS (4-< ONCE/WEEK)
99 REPORTS5	0.97297	GIVE REPORTS (5-NEVER)
100 EXPERMT1	0.80530	DO EXPERIMENTS (1-EVERYDAY)
101 EXPERMT2	0.92282	DO EXPERIMENTS (2-SEVERAL/WEEK)
102 EXPERMT3	0.96484	DO EXPERIMENTS (3-ONCE/WEEK)
103 EXPERMT4	0.93911	DO EXPERIMENTS (4-< ONCE/WEEK)
104 EXPERMT5	0.86050	DO EXPERIMENTS (5-NEVER)
105 TCHLECT1	0.97778	TEACHER LECTURES (1-EVERYDAY)
106 TCHLECT2	0.98532	TEACHER LECTURES (2-SEVERAL/WEEK)
107 TCHLECT3	0.98885	TEACHER LECTURES (3-ONCE/WEEK)
108 TCHLECT4	0.98659	TEACHER LECTURES (4-< ONCE/WEEK)
109 TCHLECT5	0.95685	TEACHER LECTURES (5-NEVER)
110 TCHDEMO1	0.95023	TEACH DEMOSTRATES (1-EVERYDAY)
111 TCHDEMO2	0.96141	TEACH DEMOSTRATES (2-SEVERAL/WEEK)
112 TCHDEMO3	0.98204	TEACH DEMOSTRATES (3-ONCE/WEEK)
113 TCHDEMO4	0.96857	TEACH DEMOSTRATES (4-< ONCE/WEEK)
114 TCHDEMO5	0.81333	TEACH DEMOSTRATES (5-NEVER)
115 TCHREAS1	0.89754	TEACH ASKS FOR REASONS (1-EVERYDAY)
116 TCHREAS2	0.95523	TEACH ASKS FOR REASONS (2-SEVERAL/WEEK)
117 TCHREAS3	0.97171	TEACH ASKS FOR REASONS (3-ONCE/WEEK)
118 TCHREAS4	0.96377	TEACH ASKS FOR REASONS (4-< ONCE/WEEK)
119 TCHREAS5	0.78509	TEACH ASKS FOR REASONS (5-NEVER)
120 WRITEXP1	0.85249	WRITE UP EXPERIMENTS (1-EVERYDAY)
121 WRITEXP2	0.96231	WRITE UP EXPERIMENTS (2-SEVERAL/WEEK)
122 WRITEXP3	0.97772	WRITE UP EXPERIMENTS (3-ONCE/WEEK)
123 WRITEXP4	0.96737	WRITE UP EXPERIMENTS (4-< ONCE/WEEK)
124 WRITEXP5	0.93836	WRITE UP EXPERIMENTS (5-NEVER)
125 GOPNION1	0.96208	GIVE OPINIONS (1-EVERYDAY)
126 GOPNION2	0.96933	GIVE OPINIONS (2-SEVERAL/WEEK)
127 GOPNION3	0.98667	GIVE OPINIONS (3-ONCE/WEEK)
128 GOPNION4	0.98516	GIVE OPINIONS (4-< ONCE/WEEK)
129 GOPNION5	0.92834	GIVE OPINIONS (5-NEVER)
130 USECOMP1	0.96015	USE COMPUTERS (1-EVERYDAY)
131 USECOMP2	0.98079	USE COMPUTERS (2-SEVERAL/WEEK)
132 USECOMP3	0.98478	USE COMPUTERS (3-ONCE/WEEK)
133 USECOMP4	0.98114	USE COMPUTERS (4-< ONCE/WEEK)
134 USECOMP5	0.97131	USE COMPUTERS (5-NEVER)
135 USEMICR1	0.98484	USED A MICROSCOPE (1-YES)
136 USEMICR2	0.98205	USED A MICROSCOPE (2-NO)

Table F-26 (continued)
Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Science Cross-sectional Conditioning Variables, Age 13/Grade 8

Contrast	Proportion of Variance	Description
137 USEBALN1	0.98700	USED A BALANCE (1-YES)
138 USEBALN2	0.98137	USED A BALANCE (2-NO)
139 USEBURN1	0.97769	USED A GAS BURNER (1-YES)
140 USEBURN2	0.98318	USED A GAS BURNER (2-NO)
141 USEWAVE1	0.97191	USED A WAVE TANK (1-YES)
142 USEWAVE2	0.97254	USED A WAVE TANK (2-NO)
143 TSCIENC1	0.86868	TEACHER: YEARS TEACHING SCIENCE (> 20)
144 TSCIENC2	0.89544	TEACHER: YEARS TEACHING SCIENCE (11-20)
145 TSCIENC3	0.93130	TEACHER: YEARS TEACHING SCIENCE (6-10)
146 TSCIENC4	0.92998	TEACHER: YEARS TEACHING SCIENCE (2-5)
147 TSCIENC5	0.88285	TEACHER: YEARS TEACHING SCIENCE (0-1)
148 TEDUCATN	0.90186	TEACHER: EDUCATION COURSES (0-7)
149 TEDUCTMS	0.82353	TEACHER: EDUCATION COURSES (MISSING)
150 TBIOLIFE	0.81469	TEACHER: BIOLOGY/LIFE SCI COURSES (0-7)
151 TBIOLFMS	0.86254	TEACHER: BIOLOGY/LIFE SCI COURSES (MISS)
152 TCHMSTRY	0.81099	TEACHER: CHEMISTRY COURSES (0-7)
153 TCHMSTMS	0.78916	TEACHER: CHEMISTRY COURSES (MISSING)
154 TPHYSICS	0.80595	TEACHER: PHYSICS COURSES (0-7)
155 TPHYSCMS	0.82959	TEACHER: PHYSICS COURSES (MISSING)
156 TEARTHSC	0.72472	TEACHER: EARTH SCI COURSES (0-7)
157 TEASCIMS	0.83687	TEACHER: EARTH SCI COURSES (MISSING)
158 TLABFC12	0.92973	TEACHER: ADEQUATE LAB FAC (1,2-AGREE)
159 TLABFAC3	0.79704	TEACHER: ADEQUATE LAB FAC (3-NO OPINION)
160 TLABFC45	0.91282	TEACHER: ADEQUATE LAB FAC (4,5-DISAGREE)
161 TINSMT12	0.92998	TEACHER: INSTRCTNL MATERL (1,2-AGREE)
162 TINSMTL3	0.78245	TEACHER: INSTRCTNL MATERL (3-NO OPINION)
163 TINSMT45	0.91447	TEACHER: INSTRCTNL MATERL (4,5-DISAGREE)
164 TTXTCU12	0.92845	TEACHER: TEXTBOOK CURRCLM (1,2-AGREE)
165 TTXTCUR3	0.91989	TEACHER: TEXTBOOK CURRCLM (3-NO OPINION)
166 TTXTCU45	0.92678	TEACHER: TEXTBOOK CURRCLM (4,5-DISAGREE)
167 TKNWFCT1	0.91703	TEACHER: EMPH KNOWING FACTS (1-HEAVY)
168 TKNWFCT2	0.94092	TEACHER: EMPH KNOWING FACTS (2-MODERATE)
169 TKNWFCT3	0.82057	TEACHER: EMPH KNOWING FACTS (3-LITTLE)
170 TKNWFCT4	0.00000	TEACHER: EMPH KNOWING FACTS (4-NONE)
171 TKEYCON1	0.95236	TEACHER: EMPH KEY CONCEPTS (1-HEAVY)
172 TKEYCON2	0.90117	TEACHER: EMPH KEY CONCEPTS (2-MODERATE)
173 TKEYCON3	0.80843	TEACHER: EMPH KEY CONCEPTS (3-LITTLE)
174 TKEYCON4	0.00000	TEACHER: EMPH KEY CONCEPTS (4-NONE)
175 TPRBSOL1	0.94007	TEACHER: EMPH PRBLM-SOLVNG (1-HEAVY)
176 TPRBSOL2	0.92270	TEACHER: EMPH PRBLM-SOLVNG (2-MODERATE)
177 TPRBOLS3	0.88324	TEACHER: EMPH PRBLM-SOLVNG (3-LITTLE)
178 TPRBOLS4	0.00000	TEACHER: EMPH PRBLM-SOLVNG (4-NONE)
179 TNATSCI1	0.88450	TEACHER: EMPH NATURE SCI (1-HEAVY)
180 TNATSCI2	0.96187	TEACHER: EMPH NATURE SCI (2-MODERATE)
181 TNATSCI3	0.91548	TEACHER: EMPH NATURE SCI (3-LITTLE)

Table F-26 (continued)
Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Science Cross-sectional Conditioning Variables, Age 13/Grade 8

Contrast	Proportion of Variance	Description
182 TNATSCI4	0.89329	TEACHER: EMPH NATURE SCI (4-NONE)
183 TLABTNQ1	0.84629	TEACHER: EMPH LAB TECHNQS (1-HEAVY)
184 TLABTNQ2	0.89685	TEACHER: EMPH LAB TECHNQS (2-MODERATE)
185 TLABTNQ3	0.84673	TEACHER: EMPH LAB TECHNQS (3-LITTLE)
186 TLABTNQ4	0.87448	TEACHER: EMPH LAB TECHNQS (4-NONE)
187 TABLCLS1	0.94767	TEACHER: ABILITY OF CLASS (1-HIGH)
188 TABLCL24	0.95763	TEACHER: ABILITY OF CLASS (2,4-AVG,MIX)
189 TABLCLS3	0.96101	TEACHER: ABILITY OF CLASS (3-LOW)
190 TCOURSE1	0.85588	TEACHER: COURSE CONTENT (1-GEN SCI)
191 TCOURSE2	0.84091	TEACHER: COURSE CONTENT (2-LIFE SCI)
192 TCOURSE3	0.88831	TEACHER: COURSE CONTENT (3-EARTH SCI)
193 TCOURSE4	0.85614	TEACHER: COURSE CONTENT (4-PHYSICAL SCI)
194 TCOURSE5	0.86188	TEACHER: COURSE CONTENT (5-INTEGRTD SCI)
195 TLECTR1	0.88077	TEACHER: LECTURE (1-ALMOST EVERYDAY)
196 TLECTR2	0.92443	TEACHER: LECTURE (2-SEVERAL/WEEK)
197 TLECTR34	0.87160	TEACHER: LECTURE (3,4-ONCE/WEEK OR LESS)
198 TLECTR5	0.95008	TEACHER: LECTURE (5-NEVER)
199 TTXTBK1	0.86868	TEACHER: READ TEXTBOOK (1-ALMOST DAILY)
200 TTXTBK2	0.94419	TEACHER: READ TEXTBOOK (2-SEVERAL/WEEK)
201 TTXTBK34	0.88765	TEACHER: READ TEXTBOOK (3,4-1/WEEK OR <)
202 TTXTBK5	0.80221	TEACHER: READ TEXTBOOK (5-NEVER)
203 TCOMPT1	0.80565	TEACHER: USE COMPUTER (1-ALMOST EVERYDAY
204 TCOMPT2	0.88134	TEACHER: USE COMPUTER (2-SEVERAL/WEEK)
205 TCOMPT34	0.81329	TEACHER: USE COMPUTER (3-ONCE/WEEK)
206 TCOMPT4	0.93391	TEACHER: USE COMPUTER (4-< ONCE/WEEK)
207 TCOMPT5	0.92975	TEACHER: USE COMPUTER (5-NEVER)
208 TSCIACTV	0.92494	TEACHER: LEVEL OF SCI ACTIVITIES (0-12)
209 THOMEWRK	0.85039	TEACHER: TIME ON HOMEWORK WEEKLY (0-4)
210 TMATCH_2	0.93889	TEACHER: MATCH STATUS (PARTIAL)
211 TMATCH_1	0.98190	TEACHER: MATCH STATUS (UNMATCHED)
212 TMTCHXR1	0.86080	TEACHER: MATCH STATUS X RACE 1
213 TMTCHXR2	0.85759	TEACHER: MATCH STATUS X RACE 2
214 TMTCHXR3	0.90012	TEACHER: MATCH STATUS X RACE 3
215 TMTCHXR4	0.88492	TEACHER: MATCH STATUS X RACE 4
216 TMTCHXR5	0.83447	TEACHER: MATCH STATUS X RACE 5
217 TMTCHXR6	0.84612	TEACHER: MATCH STATUS X RACE 6
218 TMTCHXR7	0.88157	TEACHER: MATCH STATUS X RACE 7
219 TMTCHXR8	0.87830	TEACHER: MATCH STATUS X RACE 8
220 TMTCHXP1	0.79918	TEACHER: MATCH STATUS X PARED 1
221 TMTCHXP2	0.82081	TEACHER: MATCH STATUS X PARED 2
222 TMTCHXP3	0.84529	TEACHER: MATCH STATUS X PARED 3
223 TMTCHXP4	0.86219	TEACHER: MATCH STATUS X PARED 4
224 TMTCHXP5	0.79810	TEACHER: MATCH STATUS X PARED 5
225 TMTCHXP6	0.82142	TEACHER: MATCH STATUS X PARED 6
226 TMTCHXP7	0.86724	TEACHER: MATCH STATUS X PARED 7

Table F-26 (continued)
Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Science Cross-sectional Conditioning Variables, Age 13/Grade 8

Contrast	Proportion of Variance	Description
227 TMTCHXP8	0.81202	TEACHER: MATCH STATUS X PARED 8
228 TMTCHXR1	0.82813	TEACHER: MATCH STATUS X REGION 1
229 TMTCHXR2	0.89698	TEACHER: MATCH STATUS X REGION 2
230 TMTCHXR3	0.83611	TEACHER: MATCH STATUS X REGION 3
231 TMTCHXR4	0.90483	TEACHER: MATCH STATUS X REGION 4
232 TMTCHXR5	0.82629	TEACHER: MATCH STATUS X REGION 5
233 TMTCHXR6	0.85505	TEACHER: MATCH STATUS X REGION 6
234 TMTCHXS1	0.92693	TEACHER: MATCH STATUS X SEX 1
235 TMTCHXS2	0.90646	TEACHER: MATCH STATUS X SEX 2
236 SEASON_W	0.82468	SEASON - WINTER

Table F-27
Estimated Effects for the Science Cross-sectional Principal Components
Age 13/Grade 8

Variable	Estimated Effect	Description
1 OVERALL	0.167709	OVERALL CONSTANT
2 PCFS1	0.014079	PRINCIPAL COMPONENT FACTOR SCORE 1
3 PCFS2	-0.034528	PRINCIPAL COMPONENT FACTOR SCORE 2
4 PCFS3	-0.017681	PRINCIPAL COMPONENT FACTOR SCORE 3
5 PCFS4	0.007923	PRINCIPAL COMPONENT FACTOR SCORE 4
6 PCFS5	-0.007797	PRINCIPAL COMPONENT FACTOR SCORE 5
7 PCFS6	-0.049994	PRINCIPAL COMPONENT FACTOR SCORE 6
8 PCFS7	-0.009763	PRINCIPAL COMPONENT FACTOR SCORE 7
9 PCFS8	-0.000189	PRINCIPAL COMPONENT FACTOR SCORE 8
10 PCFS9	0.022723	PRINCIPAL COMPONENT FACTOR SCORE 9
11 PCFS10	0.008906	PRINCIPAL COMPONENT FACTOR SCORE 10
12 PCFS11	0.007673	PRINCIPAL COMPONENT FACTOR SCORE 11
13 PCFS12	-0.017949	PRINCIPAL COMPONENT FACTOR SCORE 12
14 PCFS13	-0.033027	PRINCIPAL COMPONENT FACTOR SCORE 13
15 PCFS14	0.007419	PRINCIPAL COMPONENT FACTOR SCORE 14
16 PCFS15	0.006905	PRINCIPAL COMPONENT FACTOR SCORE 15
17 PCFS16	0.001552	PRINCIPAL COMPONENT FACTOR SCORE 16
18 PCFS17	0.008132	PRINCIPAL COMPONENT FACTOR SCORE 17
19 PCFS18	-0.010844	PRINCIPAL COMPONENT FACTOR SCORE 18
20 PCFS19	-0.025471	PRINCIPAL COMPONENT FACTOR SCORE 19
21 PCFS20	-0.017030	PRINCIPAL COMPONENT FACTOR SCORE 20
22 PCFS21	0.006209	PRINCIPAL COMPONENT FACTOR SCORE 21
23 PCFS22	0.026134	PRINCIPAL COMPONENT FACTOR SCORE 22
24 PCFS23	-0.012037	PRINCIPAL COMPONENT FACTOR SCORE 23
25 PCFS24	0.000884	PRINCIPAL COMPONENT FACTOR SCORE 24
26 PCFS25	0.025357	PRINCIPAL COMPONENT FACTOR SCORE 25
27 PCFS26	-0.022887	PRINCIPAL COMPONENT FACTOR SCORE 26
28 PCFS27	-0.019695	PRINCIPAL COMPONENT FACTOR SCORE 27
29 PCFS28	-0.006665	PRINCIPAL COMPONENT FACTOR SCORE 28
30 PCFS29	0.041427	PRINCIPAL COMPONENT FACTOR SCORE 29
31 PCFS30	0.016722	PRINCIPAL COMPONENT FACTOR SCORE 30
32 PCFS31	-0.012365	PRINCIPAL COMPONENT FACTOR SCORE 31
33 PCFS32	0.011656	PRINCIPAL COMPONENT FACTOR SCORE 32
34 PCFS33	-0.000080	PRINCIPAL COMPONENT FACTOR SCORE 33
35 PCFS34	-0.010805	PRINCIPAL COMPONENT FACTOR SCORE 34
36 PCFS35	0.000305	PRINCIPAL COMPONENT FACTOR SCORE 35
37 PCFS36	0.017420	PRINCIPAL COMPONENT FACTOR SCORE 36
38 PCFS37	-0.013957	PRINCIPAL COMPONENT FACTOR SCORE 37
39 PCFS38	0.012226	PRINCIPAL COMPONENT FACTOR SCORE 38
40 PCFS39	-0.007891	PRINCIPAL COMPONENT FACTOR SCORE 39
41 PCFS40	0.007252	PRINCIPAL COMPONENT FACTOR SCORE 40
42 PCFS41	-0.021926	PRINCIPAL COMPONENT FACTOR SCORE 41
43 PCFS42	0.020672	PRINCIPAL COMPONENT FACTOR SCORE 42
44 PCFS43	0.014674	PRINCIPAL COMPONENT FACTOR SCORE 43
45 PCFS44	-0.005396	PRINCIPAL COMPONENT FACTOR SCORE 44
46 PCFS45	-0.018066	PRINCIPAL COMPONENT FACTOR SCORE 45
47 PCFS46	-0.012108	PRINCIPAL COMPONENT FACTOR SCORE 46

Table F-27 (continued)
Estimated Effects for the Science Cross-sectional Principal Components
Age 13/Grade 8

	Variable	Estimated Effect	Description
48	PCFS47	0.009339	PRINCIPAL COMPONENT FACTOR SCORE 47
49	PCFS48	-0.001899	PRINCIPAL COMPONENT FACTOR SCORE 48
50	PCFS49	-0.005009	PRINCIPAL COMPONENT FACTOR SCORE 49
51	PCFS50	0.008731	PRINCIPAL COMPONENT FACTOR SCORE 50
52	PCFS51	-0.011587	PRINCIPAL COMPONENT FACTOR SCORE 51
53	PCFS52	-0.001815	PRINCIPAL COMPONENT FACTOR SCORE 52
54	PCFS53	0.005646	PRINCIPAL COMPONENT FACTOR SCORE 53
55	PCFS54	0.001355	PRINCIPAL COMPONENT FACTOR SCORE 54
56	PCFS55	-0.001555	PRINCIPAL COMPONENT FACTOR SCORE 55
57	PCFS56	-0.007728	PRINCIPAL COMPONENT FACTOR SCORE 56
58	PCFS57	0.012002	PRINCIPAL COMPONENT FACTOR SCORE 57
59	PCFS58	0.011291	PRINCIPAL COMPONENT FACTOR SCORE 58
60	PCFS59	0.018247	PRINCIPAL COMPONENT FACTOR SCORE 59
61	PCFS60	-0.028672	PRINCIPAL COMPONENT FACTOR SCORE 60
62	PCFS61	-0.003471	PRINCIPAL COMPONENT FACTOR SCORE 61
63	PCFS62	0.005554	PRINCIPAL COMPONENT FACTOR SCORE 62
64	PCFS63	-0.004587	PRINCIPAL COMPONENT FACTOR SCORE 63
65	PCFS64	0.012795	PRINCIPAL COMPONENT FACTOR SCORE 64
66	PCFS65	0.022383	PRINCIPAL COMPONENT FACTOR SCORE 65
67	PCFS66	-0.005173	PRINCIPAL COMPONENT FACTOR SCORE 66
68	PCFS67	0.022682	PRINCIPAL COMPONENT FACTOR SCORE 67
69	PCFS68	-0.026825	PRINCIPAL COMPONENT FACTOR SCORE 68
70	PCFS69	-0.019696	PRINCIPAL COMPONENT FACTOR SCORE 69
71	PCFS70	-0.002101	PRINCIPAL COMPONENT FACTOR SCORE 70
72	PCFS71	0.009111	PRINCIPAL COMPONENT FACTOR SCORE 71
73	PCFS72	-0.011481	PRINCIPAL COMPONENT FACTOR SCORE 72
74	PCFS73	-0.014638	PRINCIPAL COMPONENT FACTOR SCORE 73
75	PCFS74	-0.009074	PRINCIPAL COMPONENT FACTOR SCORE 74
76	PCFS75	0.015176	PRINCIPAL COMPONENT FACTOR SCORE 75
77	PCFS76	0.009996	PRINCIPAL COMPONENT FACTOR SCORE 76
78	PCFS77	0.009810	PRINCIPAL COMPONENT FACTOR SCORE 77
79	PCFS78	0.007337	PRINCIPAL COMPONENT FACTOR SCORE 78
80	PCFS79	-0.016363	PRINCIPAL COMPONENT FACTOR SCORE 79
81	PCFS80	-0.000746	PRINCIPAL COMPONENT FACTOR SCORE 80
82	PCFS81	0.009227	PRINCIPAL COMPONENT FACTOR SCORE 81
83	PCFS82	-0.039892	PRINCIPAL COMPONENT FACTOR SCORE 82
84	PCFS83	0.014571	PRINCIPAL COMPONENT FACTOR SCORE 83
85	PCFS84	-0.013364	PRINCIPAL COMPONENT FACTOR SCORE 84
86	PCFS85	0.020198	PRINCIPAL COMPONENT FACTOR SCORE 85
87	PCFS86	0.014635	PRINCIPAL COMPONENT FACTOR SCORE 86
88	PCFS87	0.002670	PRINCIPAL COMPONENT FACTOR SCORE 87
89	PCFS88	0.015122	PRINCIPAL COMPONENT FACTOR SCORE 88
90	PCFS89	0.018872	PRINCIPAL COMPONENT FACTOR SCORE 89
91	PCFS90	0.011707	PRINCIPAL COMPONENT FACTOR SCORE 90
92	PCFS91	-0.039523	PRINCIPAL COMPONENT FACTOR SCORE 91
93	PCFS92	0.027939	PRINCIPAL COMPONENT FACTOR SCORE 92
94	PCFS93	0.004601	PRINCIPAL COMPONENT FACTOR SCORE 93

Table F-27 (continued)
Estimated Effects for the Science Cross-sectional Principal Components
Age 13/Grade 8

	Variable	Estimated Effect	Description
95	PCFS94	0.011615	PRINCIPAL COMPONENT FACTOR SCORE 94
96	PCFS95	0.009834	PRINCIPAL COMPONENT FACTOR SCORE 95
97	PCFS96	-0.009260	PRINCIPAL COMPONENT FACTOR SCORE 96
98	PCFS97	-0.000884	PRINCIPAL COMPONENT FACTOR SCORE 97
99	PCFS98	-0.004893	PRINCIPAL COMPONENT FACTOR SCORE 98
100	PCFS99	-0.006339	PRINCIPAL COMPONENT FACTOR SCORE 99
101	PCFS100	0.029626	PRINCIPAL COMPONENT FACTOR SCORE 100
102	PCFS101	0.005263	PRINCIPAL COMPONENT FACTOR SCORE 101
103	PCFS102	-0.015303	PRINCIPAL COMPONENT FACTOR SCORE 102
104	PCFS103	0.000310	PRINCIPAL COMPONENT FACTOR SCORE 103
105	PCFS104	0.003574	PRINCIPAL COMPONENT FACTOR SCORE 104
106	PCFS105	0.030971	PRINCIPAL COMPONENT FACTOR SCORE 105
107	PCFS106	-0.032911	PRINCIPAL COMPONENT FACTOR SCORE 106
108	PCFS107	0.011602	PRINCIPAL COMPONENT FACTOR SCORE 107
109	PCFS108	0.016957	PRINCIPAL COMPONENT FACTOR SCORE 108
110	PCFS109	0.0[illegible]9925	PRINCIPAL COMPONENT FACTOR SCORE 109
111	PCFS110	-0.007255	PRINCIPAL COMPONENT FACTOR SCORE 110
112	PCFS111	0.019044	PRINCIPAL COMPONENT FACTOR SCORE 111
113	PCFS112	-0.009121	PRINCIPAL COMPONENT FACTOR SCORE 112
114	PCFS113	0.029801	PRINCIPAL COMPONENT FACTOR SCORE 113
115	PCFS114	-0.024938	PRINCIPAL COMPONENT FACTOR SCORE 114
116	PCFS115	-0.004789	PRINCIPAL COMPONENT FACTOR SCORE 115
117	PCFS116	0.010002	PRINCIPAL COMPONENT FACTOR SCORE 116
118	PCFS117	-0.038990	PRINCIPAL COMPONENT FACTOR SCORE 117
119	PCFS118	-0.007597	PRINCIPAL COMPONENT FACTOR SCORE 118
120	PCFS119	-0.030854	PRINCIPAL COMPONENT FACTOR SCORE 119
121	PCFS120	0.017715	PRINCIPAL COMPONENT FACTOR SCORE 120
122	PCFS121	-0.004867	PRINCIPAL COMPONENT FACTOR SCORE 121
123	PCFS122	0.009752	PRINCIPAL COMPONENT FACTOR SCORE 122
124	PCFS123	0.014848	PRINCIPAL COMPONENT FACTOR SCORE 123
125	PCFS124	-0.046340	PRINCIPAL COMPONENT FACTOR SCORE 124
126	PCFS125	0.018725	PRINCIPAL COMPONENT FACTOR SCORE 125
127	PCFS126	-0.001556	PRINCIPAL COMPONENT FACTOR SCORE 126
128	PCFS127	-0.008269	PRINCIPAL COMPONENT FACTOR SCORE 127
129	PCFS128	0.021616	PRINCIPAL COMPONENT FACTOR SCORE 128
130	PCFS129	0.002474	PRINCIPAL COMPONENT FACTOR SCORE 129
131	PCFS130	-0.004852	PRINCIPAL COMPONENT FACTOR SCORE 130

Table F-28
Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Science Cross-sectional Conditioning Variables, Age 17/Grade 12

Contrast	Proportion of Variance	Description
2 GENDER2	0.67661	SEX (FEMALE)
3 ETHNIC2	0.73817	DERIVED RACE (BLACK)
4 ETHNIC3	0.80671	DERIVED RACE (HISPANIC)
5 ETHNIC4	0.92650	DERIVED RACE (ASIAN)
6 STOC3	0.84728	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7 STOC147	0.84298	SIZE AND TYPE OF COMMUNITY (NON-HI+LOW)
8 REGION2	0.89303	REGION (SOUTHEAST)
9 REGION3	0.88777	REGION (CENTRAL)
10 REGION4	0.88463	REGION (WEST)
11 PARED2	0.95559	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12 PARED3	0.96588	PARENTS EDUCATION (POST HIGH SCHOOL)
13 PARED4	0.94982	PARENTS EDUCATION (COLLEGE GRAD)
14 PARED_	0.97027	PARENTS EDUCATION (MISSING,I DONT KNOW)
15 HOMEITM2	0.88426	3 ITEMS IN THE HOME
16 HOMEITM3	0.86783	4 ITEMS IN THE HOME
17 TVWATCHL	0.97379	# HOURS OF DAILY TV VIEWING (LINEAR)
18 TVWATCHQ	0.97502	# HOURS OF DAILY TV VIEWING (QUADRATIC)
19 LANGHM23	0.77489	LANG OTHER THAN ENG AT HOME(SOME/ALWAYS)
20 HW-NO	0.97921	HOMEWORK (NONE ASSIGNED)
21 HW-YES	0.98093	HOMEWORK (YES, ASSIGNED)
22 HWLINEAR	0.93775	HOMEWORK (LINEAR)
23 HW:QUAD	0.91682	HOMEWORK (QUADRATIC)
24 PCTLUNCH	0.79048	PERCENT IN LUNCH PROGRAM (LINEAR)
25 %LUNCH M	0.87356	PERCENT IN LUNCH PROGRAM - MISSING
26 PCTWHT1	0.79532	PERCENT WHITE IN SCHOOL (0-49 %) WHT MIN
27 PCTWHT2	0.91160	PERCENT WHITE IN SCHOOL (50-79%) INTGRAT
28 AGE/GRD2	0.97289	AGE X GRADE: MODAL AGE, < MODAL GRADE
29 AGE/GRD3	0.98595	AGE X GRADE: MODAL AGE, MODAL GRADE,MISS
30 AGE/GRD4	0.00000	AGE X GRADE: MODAL AGE, > MODAL GRADE
31 AGE/GRD5	0.99020	AGE X GRADE: > MODAL AGE, MODAL GRADE
32 SCHTYPE	0.79286	SCHOOL TYPE: NON-PUBLIC
33 HW HELP	0.95056	SOMEONE AT HOME HELPS WITH HW:>ONCE WEEK
34 SINGLEP1	0.77927	HOW MANY PARENTS AT HOME? (BOTH)
35 MOMHOME1	0.80051	DOES YOUR MOTHER LIVE AT HOME? (YES)
36 MOMWORKY	0.70276	DOES YOUR MOTHER WORK FOR PAY? (YES)
37 PGS RD14	0.80709	PAGES/DAY READ FOR SCHOOL (6 OR MORE)
38 PGS RD13	0.81513	PAGES/DAY READ FOR SCHOOL (11 OR MORE)
39 SCHMSS12	0.91607	SCHOOL DAYS MISSED LAST MONTH (0-2 DAYS)
40 HS PROG2	0.63719	HIGH SCHOOL PROGRAM (2-COLLEGE PREP)
41 HS PROG3	0.86861	HIGH SCHOOL PROGRAM (3-VOCATNL,TECHNCL)
42 POST HS2	0.88552	POST-SECONDARY PLANS (2-TWO-YR COLLEGE)
43 POST HS3	0.78962	POST-SECONDARY PLANS (3-FOUR-YR COLLEGE)
44 BEHAVRNM	0.82586	RULES OF BEHAVIOR ARE STRICT (NON-MISSG)
45 BEHAVR-L	0.81084	RULES OF BEHAVIOR ARE STRICT (LINEAR)
46 SAFETYNM	0.86776	DON'T FEEL SAFE AT SCHOOL (NON-MISSG)

Table F-28 (continued)
Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Science Cross-sectional Conditioning Variables, Age 17/Grade 12

Contrast	Proportion of Variance	Description
47 SAFETY-L	0.65078	DON'T FEEL SAFE AT SCHOOL (LINEAR)
48 DISRPTNM	0.80340	STUDENTS OFTEN DISRUPT CLASS (NON-MISSG)
49 DISRPT-L	0.70861	STUDENTS OFTEN DISRUPT CLASS (LINEAR)
50 SEMENGNM	0.86090	SEMESTERS OF ENGLSH/LIT/WRITE (NON-MISS)
51 SEMENG-L	0.86708	SEMESTERS OF ENGLSH/LIT/WRITE (LINEAR)
52 SEMMTHNM	0.91156	SEMESTERS OF MATHEMATICS (NON-MISSING)
53 SEMMTH-L	0.88833	SEMESTERS OF MATHEMATICS (LINEAR)
54 SEMSCINM	0.91560	SEMESTERS OF SCIENCE (NON-MISSING)
55 SEMSCI-L	0.87514	SEMESTERS OF SCIENCE (LINEAR)
56 SEMHISNM	0.92095	SEMESTERS OF HIST/SS/GEOGRPHY (NON-MISS)
57 SEMHIS-L	0.82942	SEMESTERS OF HIST/SS/GEOGRPHY (LINEAR)
58 SEMFLGNM	0.92599	SEMESTERS OF FOREIGN LANGUAGE (NON-MISS)
59 SEMFLG-L	0.69921	SEMESTERS OF FOREIGN LANGUAGE (LINEAR)
60 SEMVOCNM	0.91416	SEMESTERS OF VOC/TCH/BUS ED (NON-MISS)
61 SEMVOC-L	0.66426	SEMESTERS OF VOC/TCH/BUS ED (LINEAR)
62 SEMARTNM	0.91157	SEMESTERS OF ART/MUSIC (NON-MISSING)
63 SEMART-L	0.70835	SEMESTERS OF ART/MUSIC (LINEAR)
64 SCHMATH	0.83169	SCHOOL LEVEL PROFICIENCY - MATH
65 SCHMATHM	0.77719	SCHOOL LEVEL PROFICIENCY MATH (MISSING)
66 SCHREAD	0.79853	SCHOOL LEVEL PROFICIENCY - READING
67 SCHREADM	0.78921	SCHOOL LEVEL PROFICIENCY READING (MISSG)
68 SCHSCI	0.83769	SCHOOL LEVEL PROFICIENCY - SCIENCE
69 SCHSCIM	0.81093	SCHOOL LEVEL PROFICIENCY SCIENCE (MISSG)
70 P/A_EXP1	0.99340	PLANT/ANIMAL EXPERIMENTS (YES)
71 P/A_EXP2	0.99311	PLANT/ANIMAL EXPERIMENTS (NO)
72 ELE_EXP1	0.98690	ELECTRICAL EXPERIMENTS (YES)
73 ELE_EXP2	0.98626	ELECTRICAL EXPERIMENTS (NO)
74 CHM_EXP1	0.98832	CHEMICAL EXPERIMENTS (YES)
75 CHM_EXP2	0.98823	CHEMICAL EXPERIMENTS (NO)
76 R/M_EXP1	0.99442	ROCK/MINERAL EXPERIMENTS (YES)
77 R/M_EXP2	0.99462	ROCK/MINERAL EXPERIMENTS (NO)
78 TEL_EXP1	0.99557	TELESCOPE EXPERIMENTS (YES)
79 TEL_EXP2	0.99590	TELESCOPE EXPERIMENTS (NO)
80 WEA_EXP1	0.99486	WEATHER EXPERIMENTS (YES)
81 WEA_EXP2	0.99486	WEATHER EXPERIMENTS (NO)
82 LIKESCI1	0.98783	LIKE SCIENCE? (YES)
83 LIKESCI2	0.98785	LIKE SCIENCE? (NO)
84 AMTSCI1	0.92822	AMOUNT OF SCIENCE (1-EVERY DA)
85 AMTSCI2	0.97984	AMOUNT OF SCIENCE (2-SEVERAL/WEEK)
86 AMTSCI3	0.99088	AMOUNT OF SCIENCE (3-ONCE/WEEK)
87 AMTSCI4	0.98880	AMOUNT OF SCIENCE (4-< ONCE/WEEK)
88 AMTSCI5	0.90727	AMOUNT OF SCIENCE (5-NEVER)
89 HW_DONE1	0.96138	AMOUNT OF HOMEWORK DONE (1-NONE)
90 HW_DONE2	0.96434	AMOUNT OF HOMEWORK DONE (2-1/2 HOUR)
91 HW_DONE3	0.97220	AMOUNT OF HOMEWORK DONE (3-ONE HOUR)

Table F-28 (continued)
Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Science Cross-sectional Conditioning Variables, Age 17/Grade 12

Contrast	Proportion of Variance	Description
92 HW_DONE4	0.97511	AMOUNT OF HOMEWORK DONE (4-TWO HOURS)
93 HW_DONE5	0.91933	AMOUNT OF HOMEWORK DONE (5-> 2 HOURS)
94 HW_DOME6	0.91793	AMOUNT OF HOMEWORK DONE (6-NO SCI CLASS)
95 WEEKLPJ1	0.98743	WEEK-LONG PROJECTS (YES)
96 WEEKLPJ2	0.98871	WEEK-LONG PROJECTS (NO)
97 READTXT1	0.95409	READ A TEXTBOOK (1-EVERYDAY)
98 READTXT2	0.95995	READ A TEXTBOOK (2-SEVERAL/WEEK)
99 READTXT3	0.98028	READ A TEXTBOOK (3-ONCE/WEEK)
100 READTXT4	0.98355	READ A TEXTBOOK (4-< ONCE/WEEK)
101 READTXT5	0.79588	READ A TEXTBOOK (5-NEVER)
102 SCINEWS1	0.92310	DISCUSS SCIENCE NEWS (1-EVERYDAY)
103 SCINEWS2	0.94733	DISCUSS SCIENCE NEWS (2-SEVERAL/WEEK)
104 SCINEWS3	0.97739	DISCUSS SCIENCE NEWS (3-ONCE/WEEK)
105 SCINEWS4	0.99045	DISCUSS SCIENCE NEWS (4-< ONCE/WEEK)
106 SCINEWS5	0.82219	DISCUSS SCIENCE NEWS (5-NEVER)
107 WORKOTH1	0.93443	WORK WITH OTHERS (1-EVERYDAY)
108 WORKOTH2	0.96382	WORK WITH OTHERS (2-SEVERAL/WEEK)
109 WORKOTH3	0.97221	WORK WITH OTHERS (3-ONCE/WEEK)
110 WORKOTH4	0.97860	WORK WITH OTHERS (4-< ONCE/WEEK)
111 WORKOTH5	0.81346	WORK WITH OTHERS (5-NEVER)
112 REPORTS1	0.78158	GIVE REPORTS (1-EVERYDAY)
113 REPORTS2	0.95899	GIVE REPORTS (2-SEVERAL/WEEK)
114 REPORTS3	0.96443	GIVE REPORTS (3-ONCE/WEEK)
115 REPORTS4	0.98986	GIVE REPORTS (4-< ONCE/WEEK)
116 REPORTS5	0.92814	GIVE REPORTS (5-NEVER)
117 EXPERMT1	0.83398	DO EXPERIMENTS (1-EVERYDAY)
118 EXPERMT2	0.94227	DO EXPERIMENTS (2-SEVERAL/WEEK)
119 EXPERMT3	0.94024	DO EXPERIMENTS (3-ONCE/WEEK)
120 EXPERMT4	0.90391	DO EXPERIMENTS (4-< ONCE/WEEK)
121 EXPERMT5	0.80561	DO EXPERIMENTS (5-NEVER)
122 TCHLECT1	0.96033	TEACHER LECTURES (1-EVERYDAY)
123 TCHLECT2	0.95915	TEACHER LECTURES (2-SEVERAL/WEEK)
124 TCHLECT3	0.97875	TEACHER LECTURES (3-ONCE/WEEK)
125 TCHLECT4	0.98001	TEACHER LECTURES (4-< ONCE/WEEK)
126 TCHLECT5	0.73671	TEACHER LECTURES (5-NEVER)
127 TCHDEMO1	0.86849	TEACH DEMOSTRATES (1-EVERYDAY)
128 TCHDEMO2	0.97429	TEACH DEMOSTRATES (2-SEVERAL/WEEK)
129 TCHDEMO3	0.96395	TEACH DEMOSTRATES (3-ONCE/WEEK)
130 TCHDEMO4	0.91207	TEACH DEMOSTRATES (4-< ONCE/WEEK)
131 TCHDEMO5	0.80847	TEACH DEMOSTRATES (5-NEVER)
132 TCHREAS1	0.80983	TEACH ASKS FOR REASONS (1-EVERYDAY)
133 TCHREAS2	0.95710	TEACH ASKS FOR REASONS (2-SEVERAL/WEEK)
134 TCHREAS3	0.94798	TEACH ASKS FOR REASONS (3-ONCE/WEEK)
135 TCHREAS4	0.87405	TEACH ASKS FOR REASONS (4-< ONCE/WEEK)
136 TCHREAS5	0.77798	TEACH ASKS FOR REASONS (5-NEVER)

Table F-28 (continued)
Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Science Cross-sectional Conditioning Variables, Age 17/Grade 12

Contrast	Proportion of Variance	Description
137 WRITEXP1	0.77412	WRITE UP EXPERIMENTS (1-EVERYDAY)
138 WRITEXP2	0.93783	WRITE UP EXPERIMENTS (2-SEVERAL/WEEK)
139 WRITEXP3	0.97191	WRITE UP EXPERIMENTS (3-ONCE/WEEK)
140 WRITEXP4	0.96724	WRITE UP EXPERIMENTS (4-< ONCE/WEEK)
141 WRITEXP5	0.85491	WRITE UP EXPERIMENTS (5-NEVER)
142 GOPNION1	0.92790	GIVE OPINIONS (1-EVERYDAY)
143 GOPNION2	0.96254	GIVE OPINIONS (2-SEVERAL/WEEK)
144 GOPNION3	0.98776	GIVE OPINIONS (3-ONCE/WEEK)
145 GOPNION4	0.98627	GIVE OPINIONS (4-< ONCE/WEEK)
146 GOPNION5	0.86398	GIVE OPINIONS (5-NEVER)
147 USECOMP1	0.91673	USE COMPUTERS (1-EVERYDAY)
148 USECOMP2	0.97720	USE COMPUTERS (2-SEVERAL/WEEK)
149 USECOMP3	0.98302	USE COMPUTERS (3-ONCE/WEEK)
150 USECOMP4	0.97706	USE COMPUTERS (4-< ONCE/WEEK)
151 USECOMP5	0.97635	USE COMPUTERS (5-NEVER)
152 USEMICR1	0.95614	USED A MICROSCOPE (1-YES)
153 USEMICR2	0.94952	USED A MICROSCOPE (2-NO)
154 USEBALN1	0.91661	USED A BALANCE (1-YES)
155 USEBALN2	0.87976	USED A BALANCE (2-NO)
156 USEBURN1	0.92483	USED A GAS BURNER (1-YES)
157 USEBURN2	0.93024	USED A GAS BURNER (2-NO)
158 USEWAVE1	0.96347	USED A WAVE TANK (1-YES)
159 USEWAVE2	0.97698	USED A WAVE TANK (2-NO)
160 STGENSC1	0.95042	HOW LONG STUDIED GEN SCI (1-> YEAR)
161 STGENSC2	0.97692	HOW LONG STUDIED GEN SCI (2-ONE YEAR)
162 STGENSC3	0.97253	HOW LONG STUDIED GEN SCI (3-HALF YEAR)
163 STGENSC4	0.96123	HOW LONG STUDIED GEN SCI (4-NOT STUDIED)
164 STGENSCO	0.87754	HOW LONG STUDIED GEN SCI (OMIT)
165 STBIOLG1	0.94230	HOW LONG STUDIED BIOLOGY (1-> YEAR)
166 STBIOLG2	0.98072	HOW LONG STUDIED BIOLOGY (2-ONE YEAR)
167 STBIOLG3	0.98712	HOW LONG STUDIED BIOLOGY (3-HALF YEAR)
168 STBIOLG4	0.95065	HOW LONG STUDIED BIOLOGY (4-NOT STUDIED)
169 STBIOLGO	0.72333	HOW LONG STUDIED BIOLOGY (OMIT)
170 STLIFSC1	0.74201	HOW LONG STUDIED LIFE SC (1-> YEAR)
171 STLIFSC2	0.96840	HOW LONG STUDIED LIFE SC (2-ONE YEAR)
172 STLIFSC3	0.96893	HOW LONG STUDIED LIFE SC (3-HALF YEAR)
173 STLIFSC4	0.96351	HOW LONG STUDIED LIFE SC (4-NOT STUDIED)
174 STLIFSCO	0.70729	HOW LONG STUDIED LIFE SC (OMIT)
175 STCHMST1	0.78202	HOW LONG STUDIED CHMSTRY (1-> YEAR)
176 STCHMST2	0.94606	HOW LONG STUDIED CHMSTRY (2-ONE YEAR)
177 STCHMST3	0.97282	HOW LONG STUDIED CHMSTRY (3-HALF YEAR)
178 STCHMST4	0.87663	HOW LONG STUDIED CHMSTRY (4-NOT STUDIED)
179 STCHMSTO	0.72072	HOW LONG STUDIED CHMSTRY (OMIT)
180 STPHYSC1	0.79811	HOW LONG STUDIED PHYSICS (1-> YEAR)
181 STPHYSC2	0.93107	HOW LONG STUDIED PHYSICS (2-ONE YEAR)

Table F-28 (continued)
Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Science Cross-sectional Conditioning Variables, Age 17/Grade 12

Contrast	Proportion of Variance	Description
182 STPHYSC3	0.97845	HOW LONG STUDIED PHYSICS (3-HALF YEAR)
183 STPHYSC4	0.94668	HOW LONG STUDIED PHYSICS (4-NOT STUDIED)
184 STPHYSCO	0.64160	HOW LONG STUDIED PHYSICS (OMIT)
185 STPHYCL1	0.82762	HOW LONG STUDIED PHY SCI (1-> YEAR)
186 STPHYCL2	0.96710	HOW LONG STUDIED PHY SCI (2-ONE YEAR)
187 STPHYCL3	0.96537	HOW LONG STUDIED PHY SCI (3-HALF YEAR)
188 STPHYCL4	0.95929	HOW LONG STUDIED PHY SCI (4-NOT STUDIED)
189 STPHYCLO	0.81666	HOW LONG STUDIED PHY SCI (OMIT)
190 STEA&SP1	0.73878	HOW LONG STUDIED EA + SP (1-> YEAR)
191 STEA&SP2	0.95826	HOW LONG STUDIED EA + SP (2-ONE YEAR)
192 STEA&SP3	0.96225	HOW LONG STUDIED EA + SP (3-HALF YEAR)
193 STEA&SP4	0.96036	HOW LONG STUDIED EA + SP (4-NOT STUDIED)
194 SEASON_W	0.98346	SEASON - WINTER

Table F-29
Estimated Effects for the Science Cross-sectional Principal Components
Age 17/Grade 12

	Variable	Estimated Effect	Description
1	OVERALL	0.712992	OVERALL CONSTANT
2	PCFS1	0.025240	PRINCIPAL COMPONENT FACTOR SCORE 1
3	PCFS2	0.003146	PRINCIPAL COMPONENT FACTOR SCORE 2
4	PCFS3	-0.029710	PRINCIPAL COMPONENT FACTOR SCORE 3
5	PCFS4	-0.038990	PRINCIPAL COMPONENT FACTOR SCORE 4
6	PCFS5	-0.029499	PRINCIPAL COMPONENT FACTOR SCORE 5
7	PCFS6	-0.004465	PRINCIPAL COMPONENT FACTOR SCORE 6
8	PCFS7	0.019515	PRINCIPAL COMPONENT FACTOR SCORE 7
9	PCFS8	0.000009	PRINCIPAL COMPONENT FACTOR SCORE 8
10	PCFS9	-0.020200	PRINCIPAL COMPONENT FACTOR SCORE 9
11	PCFS10	0.003944	PRINCIPAL COMPONENT FACTOR SCORE 10
12	PCFS11	-0.018700	PRINCIPAL COMPONENT FACTOR SCORE 11
13	PCFS12	0.001974	PRINCIPAL COMPONENT FACTOR SCORE 12
14	PCFS13	-0.009639	PRINCIPAL COMPONENT FACTOR SCORE 13
15	PCFS14	-0.000783	PRINCIPAL COMPONENT FACTOR SCORE 14
16	PCFS15	-0.032860	PRINCIPAL COMPONENT FACTOR SCORE 15
17	PCFS16	0.030739	PRINCIPAL COMPONENT FACTOR SCORE 16
18	PCFS17	0.011110	PRINCIPAL COMPONENT FACTOR SCORE 17
19	PCFS18	-0.010807	PRINCIPAL COMPONENT FACTOR SCORE 18
20	PCFS19	0.009152	PRINCIPAL COMPONENT FACTOR SCORE 19
21	PCFS20	0.028987	PRINCIPAL COMPONENT FACTOR SCORE 20
22	PCFS21	0.017898	PRINCIPAL COMPONENT FACTOR SCORE 21
23	PCFS22	0.027867	PRINCIPAL COMPONENT FACTOR SCORE 22
24	PCFS23	0.009066	PRINCIPAL COMPONENT FACTOR SCORE 23
25	PCFS24	-0.037294	PRINCIPAL COMPONENT FACTOR SCORE 24
26	PCFS25	0.010119	PRINCIPAL COMPONENT FACTOR SCORE 25
27	PCFS26	-0.014081	PRINCIPAL COMPONENT FACTOR SCORE 26
28	PCFS27	-0.011895	PRINCIPAL COMPONENT FACTOR SCORE 27
29	PCFS28	-0.003361	PRINCIPAL COMPONENT FACTOR SCORE 28
30	PCFS29	0.009724	PRINCIPAL COMPONENT FACTOR SCORE 29
31	PCFS30	-0.018732	PRINCIPAL COMPONENT FACTOR SCORE 30
32	PCFS31	0.006991	PRINCIPAL COMPONENT FACTOR SCORE 31
33	PCFS32	-0.027065	PRINCIPAL COMPONENT FACTOR SCORE 32
34	PCFS33	-0.007743	PRINCIPAL COMPONENT FACTOR SCORE 33
35	PCFS34	-0.023205	PRINCIPAL COMPONENT FACTOR SCORE 34
36	PCFS35	-0.006822	PRINCIPAL COMPONENT FACTOR SCORE 35
37	PCFS36	-0.006725	PRINCIPAL COMPONENT FACTOR SCORE 36
38	PCFS37	[illegible].019625	PRINCIPAL COMPONENT FACTOR SCORE 37
39	PCFS38	-0.001285	PRINCIPAL COMPONENT FACTOR SCORE 38
40	PCFS39	0.018074	PRINCIPAL COMPONENT FACTOR SCORE 39
41	PCFS40	-0.006875	PRINCIPAL COMPONENT FACTOR SCORE 40
42	PCFS41	0.001957	PRINCIPAL COMPONENT FACTOR SCORE 41
43	PCFS42	-0.012805	PRINCIPAL COMPONENT FACTOR SCORE 42
44	PCFS43	-0.001516	PRINCIPAL COMPONENT FACTOR SCORE 43
45	PCFS44	-0.002353	PRINCIPAL COMPONENT FACTOR SCORE 44
46	PCFS45	-0.006091	PRINCIPAL COMPONENT FACTOR SCORE 45
47	PCFS46	-0.000800	PRINCIPAL COMPONENT FACTOR SCORE 46

Table F-29 (continued)
Estimated Effects for the Science Cross-sectional Principal Components
Age 17/Grade 12

	Variable	Estimated Effect	Description
48	PCFS47	-0.013240	PRINCIPAL COMPONENT FACTOR SCORE 47
49	PCFS48	0.015549	PRINCIPAL COMPONENT FACTOR SCORE 48
50	PCFS49	-0.002205	PRINCIPAL COMPONENT FACTOR SCORE 49
51	PCFS50	0.001535	PRINCIPAL COMPONENT FACTOR SCORE 50
52	PCFS51	0.000760	PRINCIPAL COMPONENT FACTOR SCORE 51
53	PCFS52	-0.025824	PRINCIPAL COMPONENT FACTOR SCORE 52
54	PCFS53	-0.004558	PRINCIPAL COMPONENT FACTOR SCORE 53
55	PCFS54	0.014542	PRINCIPAL COMPONENT FACTOR SCORE 54
56	PCFS55	-0.010915	PRINCIPAL COMPONENT FACTOR SCORE 55
57	PCFS56	0.002396	PRINCIPAL COMPONENT FACTOR SCORE 56
58	PCFS57	0.003388	PRINCIPAL COMPONENT FACTOR SCORE 57
59	PCFS58	0.001202	PRINCIPAL COMPONENT FACTOR SCORE 58
60	PCFS59	-0.007285	PRINCIPAL COMPONENT FACTOR SCORE 59
61	PCFS60	0.026053	PRINCIPAL COMPONENT FACTOR SCORE 60
62	PCFS61	-0.028354	PRINCIPAL COMPONENT FACTOR SCORE 61
63	PCFS62	0.006782	PRINCIPAL COMPONENT FACTOR SCORE 62
64	PCFS63	0.021700	PRINCIPAL COMPONENT FACTOR SCORE 63
65	PCFS64	-0.004971	PRINCIPAL COMPONENT FACTOR SCORE 64
66	PCFS65	-0.017997	PRINCIPAL COMPONENT FACTOR SCORE 65
67	PCFS66	0.005981	PRINCIPAL COMPONENT FACTOR SCORE 66
68	PCFS67	0.016828	PRINCIPAL COMPONENT FACTOR SCORE 67
69	PCFS68	0.008890	PRINCIPAL COMPONENT FACTOR SCORE 68
70	PCFS69	0.006846	PRINCIPAL COMPONENT FACTOR SCORE 69
71	PCFS70	-0.011389	PRINCIPAL COMPONENT FACTOR SCORE 70
72	PCFS71	0.006649	PRINCIPAL COMPONENT FACTOR SCORE 71
73	PCFS72	0.010418	PRINCIPAL COMPONENT FACTOR SCORE 72
74	PCFS73	-0.011148	PRINCIPAL COMPONENT FACTOR SCORE 73
75	PCFS74	0.001499	PRINCIPAL COMPONENT FACTOR SCORE 74
76	PCFS75	0.001252	PRINCIPAL COMPONENT FACTOR SCORE 75
77	PCFS76	0.001550	PRINCIPAL COMPONENT FACTOR SCORE 76
78	PCFS77	0.002894	PRINCIPAL COMPONENT FACTOR SCORE 77
79	PCFS78	0.015086	PRINCIPAL COMPONENT FACTOR SCORE 78
80	PCFS79	-0.001377	PRINCIPAL COMPONENT FACTOR SCORE 79
81	PCFS80	-0.005634	PRINCIPAL COMPONENT FACTOR SCORE 80
82	PCFS81	-0.042789	PRINCIPAL COMPONENT FACTOR SCORE 81
83	PCFS82	-0.020364	PRINCIPAL COMPONENT FACTOR SCORE 82
84	PCFS83	-0.003803	PRINCIPAL COMPONENT FACTOR SCORE 83
85	PCFS84	-0.008014	PRINCIPAL COMPONENT FACTOR SCORE 84
86	PCFS85	0.043606	PRINCIPAL COMPONENT FACTOR SCORE 85
87	PCFS86	0.013858	PRINCIPAL COMPONENT FACTOR SCORE 86
88	PCFS87	0.009677	PRINCIPAL COMPONENT FACTOR SCORE 87
89	PCFS88	-0.023469	PRINCIPAL COMPONENT FACTOR SCORE 88
90	PCFS89	0.004243	PRINCIPAL COMPONENT FACTOR SCORE 89
91	PCFS90	-0.022804	PRINCIPAL COMPONENT FACTOR SCORE 90
92	PCFS91	-0.007189	PRINCIPAL COMPONENT FACTOR SCORE 91
93	PCFS92	-0.001420	PRINCIPAL COMPONENT FACTOR SCORE 92
94	PCFS93	0.019349	PRINCIPAL COMPONENT FACTOR SCORE 93

Table F-29 (continued)
Estimated Effects for the Science Cross-sectional Principal Components
Age 17/Grade 12

	Variable	Estimated Effect	Description
95	PCFS94	-0.035030	PRINCIPAL COMPONENT FACTOR SCORE 94
96	PCFS95	0.009696	PRINCIPAL COMPONENT FACTOR SCORE 95
97	PCFS96	0.013114	PRINCIPAL COMPONENT FACTOR SCORE 96
98	PCFS97	-0.019715	PRINCIPAL COMPONENT FACTOR SCORE 97
99	PCFS98	-0.004725	PRINCIPAL COMPONENT FACTOR SCORE 98
100	PCFS99	0.002334	PRINCIPAL COMPONENT FACTOR SCORE 99
101	PCFS100	0.007091	PRINCIPAL COMPONENT FACTOR SCORE 100
102	PCFS101	-0.014796	PRINCIPAL COMPONENT FACTOR SCORE 101
103	PCFS102	0.038498	PRINCIPAL COMPONENT FACTOR SCORE 102
104	PCFS103	-0.010057	PRINCIPAL COMPONENT FACTOR SCORE 103
105	PCFS104	-0.021958	PRINCIPAL COMPONENT FACTOR SCORE 104
106	PCFS105	-0.018702	PRINCIPAL COMPONENT FACTOR SCORE 105
107	PCFS106	0.015814	PRINCIPAL COMPONENT FACTOR SCORE 106
108	PCFS107	-0.000235	PRINCIPAL COMPONENT FACTOR SCORE 107
109	PCFS108	0.035230	PRINCIPAL COMPONENT FACTOR SCORE 108
110	PCFS109	0.000395	PRINCIPAL COMPONENT FACTOR SCORE 109

Table F-30
Estimated Effects for Science Trend Conditioning Variable Contrasts, Age 9

Contrast	Estimated Effect	Description
1 OVERALL	0.113191	OVERALL CONSTANT
2 GENDER2	-0.149142	SEX (FEMALE)
3 ETHNIC2	-0.683258	OBSERVED ETHNICITY (BLACK)
4 ETHNIC3	-0.302134	OBSERVED ETHNICITY (HISPANIC)
5 ETHNIC4	-0.009753	OBSERVED ETHNICITY (ASIAN)
6 STOC2	-0.077940	SIZE AND TYPE OF COMMUNITY (LOW METRO)
7 STOC3	0.032611	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
8 REGION2	-0.021163	REGION (SOUTHEAST)
9 REGION3	0.081375	REGION (CENTRAL)
10 REGION4	0.054434	REGION (WEST)
11 PARED2	0.192453	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12 PARED3	0.437887	PARENTS EDUCATION (POST HIGH SCHOOL)
13 PARED4	0.370268	PARENTS EDUCATION (COLLEGE GRAD)
14 PARED_	0.166730	PARENTS EDUCATION (MISSING, I DON'T KNOW
15 MODLGRD1	-0.541046	< MODAL GRADE
16 MODLGRD2	0.039475	> MODAL GRADE
17 HOMEITM2	0.222198	3 ITEMS IN THE HOME
18 HOMEITM3	0.328705	4 ITEMS IN THE HOME
19 RAC/SEX1	0.285875	BLACK, FEMALE
20 RAC/SEX2	-0.067464	HISPANIC, FEMALE
21 RAC/SEX3	0.252980	ASIAN AMERICAN, FEMALE
22 RAC/PED1	0.023034	BLACK, HIGH SCHOOL GRAD
23 RAC/PED2	-0.225184	BLACK, POST HIGH SCHOOL
24 RAC/PED3	-0.247139	BLACK, COLLEGE GRAD
25 RAC/PED4	-0.274362	BLACK, MISSING
26 RAC/PED5	-0.045121	HISPANIC, HIGH SCHOOL GRAD
27 RAC/PED6	-0.407353	HISPANIC, POST HIGH SCHOOL
28 RAC/PED7	-0.188681	HISPANIC, COLLEGE GRAD
29 RAC/PED8	-0.075472	HISPANIC, MISSING
30 RAC/PED9	-0.613567	ASIAN AMERICAN, HIGH SCHOOL GRAD
31 RAC/PE10	-0.517159	ASIAN AMERICAN, POST HIGH SCHOOL
32 RAC/PE11	-0.363563	ASIAN AMERICAN, COLLEGE GRAD
33 RAC/PE12	-0.808073	ASIAN AMERICAN, MISSING
34 SCHTYPE	0.036300	SCHOOL (NONPUBLIC)
35 LANGHM1	-0.039005	OTHER THAN ENGLISH AT HOME (SOMETIMES)
36 LANGHM2	-0.342507	OTHER THAN ENGLISH AT HOME (ALWAYS)
37 RAC/LNG1	0.183358	BLACK, ALWAYS OTHER THAN ENGLISH
38 RAC/LNG2	0.096906	BLACK, SOMETIMES OTHER THAN ENGLISH
39 RAC/LNG3	0.220563	HISPANIC, ALWAYS OTHER THAN ENGLISH
40 RAC/LNG4	0.070651	HISPANIC, SOMETIMES OTHER THAN ENGLISH
41 RAC/LNG5	0.116793	ASIAN AM, ALWAYS OTHER THAN ENGLISH
42 RAC/LNG6	0.323559	ASIAN AM, SOMETIMES OTHER THAN ENGLISH
43 DRACE2	-0.086664	DERIVED RACE (BLACK)
44 DRACE3	-0.242782	DERIVED RACE (HISPANIC)
45 DRACE4	0.042801	DERIVED RACE (ASIAN AMERICAN)

Table F-31
Estimated Effects for Science Trend Conditioning Variable Contrasts, Age 13

Contrast	Estimated Effect	Description
1 OVERALL	-0.050678	OVERALL CONSTANT
2 GENDER2	-0.263400	SEX (FEMALE)
3 ETHNIC2	-0.489860	OBSERVED ETHNICITY (BLACK)
4 ETHNIC3	-0.113651	OBSERVED ETHNICITY (HISPANIC)
5 ETHNIC4	-0.980445	OBSERVED ETHNICITY (ASIAN)
6 STOC2	-0.357321	SIZE AND TYPE OF COMMUNITY (LOW METRO)
7 STOC3	0.076734	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
8 REGION2	-0.093112	REGION (SOUTHEAST)
9 REGION3	0.023857	REGION (CENTRAL)
10 REGION4	0.008046	REGION (WEST)
11 PARED2	0.101136	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12 PARED3	0.370100	PARENTS EDUCATION (POST HIGH SCHOOL)
13 PARED4	0.479707	PARENTS EDUCATION (COLLEGE GRAD)
14 PARED_	-0.178175	PARENTS EDUCATION (MISSING, I DON'T KNOW
15 MODLGRD1	-0.480191	< MODAL GRADE
16 MODLGRD2	0.411718	> MODAL GRADE
17 HOMEITM2	0.255485	3 ITEMS IN THE HOME
18 HOMEITM3	0.380117	4 ITEMS IN THE HOME
19 RAC/SEX1	0.212494	BLACK, FEMALE
20 RAC/SEX2	-0.042517	HISPANIC, FEMALE
21 RAC/SEX3	0.007972	ASIAN AMERICAN, FEMALE
22 RAC/PED1	-0.059142	BLACK, HIGH SCHOOL GRAD
23 RAC/PED2	-0.065298	BLACK, POST HIGH SCHOOL
24 RAC/PED3	-0.235725	BLACK, COLLEGE GRAD
25 RAC/PED4	-0.080955	BLACK, MISSING
26 RAC/PED5	-0.151840	HISPANIC, HIGH SCHOOL GRAD
27 RAC/PED6	0.129267	HISPANIC, POST HIGH SCHOOL
28 RAC/PED7	-0.115026	HISPANIC, COLLEGE GRAD
29 RAC/PED8	-0.040835	HISPANIC, MISSING
30 RAC/PED9	0.009591	ASIAN AMERICAN, HIGH SCHOOL GRAD
31 RAC/PE10	0.683266	ASIAN AMERICAN, POST HIGH SCHOOL
32 RAC/PE11	0.663879	ASIAN AMERICAN, COLLEGE GRAD
33 RAC/PE12	0.605546	ASIAN AMERICAN, MISSING
34 SCHTYPE	0.097878	SCHOOL (NONPUBLIC)
35 LANGHM1	0.165151	OTHER THAN ENGLISH AT HOME (SOMETIMES)
36 LANGHM2	-0.207535	OTHER THAN ENGLISH AT HOME (ALWAYS)
37 RAC/LNG1	0.066890	BLACK, ALWAYS OTHER THAN ENGLISH
38 RAC/LNG2	0.027021	BLACK, SOMETIMES OTHER THAN ENGLISH
39 RAC/LNG3	0.169179	HISPANIC, ALWAYS OTHER THAN ENGLISH
40 RAC/LNG4	0.094049	HISPANIC, SOMETIMES OTHER THAN ENGLISH
41 RAC/LNG5	0.382125	ASIAN AM, ALWAYS OTHER THAN ENGLISH
42 RAC/LNG6	-0.182784	ASIAN AM, SOMETIMES OTHER THAN ENGLISH
43 DRACE2	-0.248317	DERIVED RACE (BLACK)
44 DRACE3	-0.422046	DERIVED RACE (HISPANIC)
45 DRACE4	0.132099	DERIVED RACE (ASIAN AMERICAN)
46 HW-NO	-0.056752	HOMEWORK (NONE ASSIGNED)
47 HW-YES	-0.092021	HOMEWORK (YES - DIDN'T DO)
48 HW-234	0.043172	HOMEWORK (1\2 HR TO 2 HOURS)

Table F-32
Estimated Effects for Science Trend Conditioning Variable Contrasts, Age 17

Contrast	Estimated Effect	Description
1 OVERALL	-0.282339	OVERALL CONSTANT
2 GENDER2	-0.333108	SEX (FEMALE)
3 ETHNIC2	-0.756363	OBSERVED ETHNICITY (BLACK)
4 ETHNIC3	-0.206462	OBSERVED ETHNICITY (HISPANIC)
5 ETHNIC4	0.047402	OBSERVED ETHNICITY (ASIAN)
6 STOC2	-0.217899	SIZE AND TYPE OF COMMUNITY (LOW METRO)
7 STOC3	0.067944	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
8 REGION2	0.003925	REGION (SOUTHEAST)
9 REGION3	0.150084	REGION (CENTRAL)
10 REGION4	0.037770	REGION (WEST)
11 PARED2	0.152283	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12 PARED3	0.310755	PARENTS EDUCATION (POST HIGH SCHOOL)
13 PARED4	0.370266	PARENTS EDUCATION (COLLEGE GRAD)
14 PARED_	-0.310489	PARENTS EDUCATION (MISSING, I DON'T KNOW
15 MODLGRD1	-0.318441	< MODAL GRADE
16 MODLGRD2	0.030781	> MODAL GRADE
17 HOMEITM2	0.120172	3 ITEMS IN THE HOME
18 HOMEITM3	0.144372	4 ITEMS IN THE HOME
19 RAC/SEX1	0.035507	BLACK, FEMALE
20 RAC/SEX2	0.092619	HISPANIC, FEMALE
21 RAC/SEX3	0.283914	ASIAN AMERICAN, FEMALE
22 RAC/PED1	-0.034058	BLACK, HIGH SCHOOL GRAD
23 RAC/PED2	0.146601	BLACK, POST HIGH SCHOOL
24 RAC/PED3	0.028159	BLACK, COLLEGE GRAD
25 RAC/PED4	0.329207	BLACK, MISSING
26 RAC/PED5	-0.465504	HISPANIC, HIGH SCHOOL GRAD
27 RAC/PED6	0.207122	HISPANIC, POST HIGH SCHOOL
28 RAC/PED7	-0.275924	HISPANIC, COLLEGE GRAD
29 RAC/PED8	-0.030261	HISPANIC, MISSING
30 RAC/PED9	-0.756093	ASIAN AMERICAN, HIGH SCHOOL GRAD
31 RAC/PE10	-0.688973	ASIAN AMERICAN, POST HIGH SCHOOL
32 RAC/PE11	-0.702763	ASIAN AMERICAN, COLLEGE GRAD
33 RAC/PE12	-0.540948	ASIAN AMERICAN, MISSING
34 SCHTYPE	0.022042	SCHOOL (NONPUBLIC)
35 LANGHM1	0.009825	OTHER THAN ENGLISH AT HOME (SOMETIMES)
36 LANGHM2	-0.371996	OTHER THAN ENGLISH AT HOME (ALWAYS)
37 RAC/LNG1	0.071430	BLACK, ALWAYS OTHER THAN ENGLISH
38 RAC/LNG2	0.099763	BLACK, SOMETIMES OTHER THAN ENGLISH
39 RAC/LNG3	0.422416	HISPANIC, ALWAYS OTHER THAN ENGLISH
40 RAC/LNG4	0.102370	HISPANIC, SOMETIMES OTHER THAN ENGLISH
41 RAC/LNG5	0.609643	ASIAN AM, ALWAYS OTHER THAN ENGLISH
42 RAC/LNG6	0.349276	ASIAN AM, SOMETIMES OTHER THAN ENGLISH
43 DRACE2	-0.206103	DERIVED RACE (BLACK)
44 DRACE3	-0.260754	DERIVED RACE (HISPANIC)
45 DRACE4	0.074552	DERIVED RACE (ASIAN AMERICAN)
46 HW-NO	-0.343765	HOMEWORK (NONE ASSIGNED)
47 HW-YES	-0.186802	HOMEWORK (YES - DIDN'T DO)
48 HW-234	0.007214	HOMEWORK (1\2 HR TO 2 HOURS)
49 NSCI-GS	-0.030378	NSCI (GENERAL SCIENCE)

Table F-32 (continued)
Estimated Effects for Science Trend Conditioning Variable Contrasts, Age 17

Contrast	Estimated Effect	Description
50 NSCI-BI	0.066096	NSCI (BIOLOGY)
51 NSCI-CH	0.586522	NSCI (CHEMISTRY)
52 NSCI-PH	0.722660	NSCI (PHYSICS)
53 HSPROG1	0.359035	HS PROGRAM (COLLEGE PREP)
54 HSPROG2	0.085950	HS PROGRAM (VOCAT/TECHN)

APPENDIX G

Open-ended Item Score Statistics

Appendix G

OPEN-ENDED ITEM SCORE STATISTICS

This appendix contains information about the open-ended items included in the scaling of data from the 1990 cross-sectional assessments of reading, mathematics, and science.

The information in the tables includes, for each subject area and age/grade, the NAEP item numbers for each of the open-ended items included in scaling, and the block that contains the item. Because each of these items was dichotomized in order to be included in the scaling process, the tables include the codes from the NAEP database that denote the range of responses and the correct responses. A portion of the responses to the open-ended items were scored twice for the purpose of examining rater reliability. For each item, the number of papers with responses that were scored a second time is listed, along with the percent agreement between raters and Cohen's Kappa based on those responses. Unlike the measures of rater agreement described in Chapter 7, these measures of rater agreement are based on the dichotomization actually used in scaling. Cohen's Kappa is a reliability estimate appropriate for items that have been dichotomized. It is described more fully in Cohen (1968).

Table G-1
Dichotomous Scoring, Percent Agreement, and Cohen's Kappa*
for the Open-ended Reading Items Used in Cross-sectional Scaling

Item	Block	Range of Response Codes	Correct Response Codes	Sample Size	Percent Agreement	Cohen's Kappa
Age 9/Grade 4						
R000807	RE	1-5	3-5	684	88	.76
R010301	RF	1-3	2-3	896	96	.93
R011509	RI	1-4	3-4	764	87	.56
Age 8/Grade 13						
R010805	RG	1-3	2-3	814	90	.81
R010007	RG	1-3	2-3	568	91	.82
Age 17/Grade 12						
R010007	RD	1-3	2-3	865	82	.79
R010805	RG	1-3	2-3	826	83	.81
R011805	RI	1-4	3-4	687	89	.89

* Cohen's Kappa is a measure of reliability that is appropriate for items that are dichotomized.

Table G-2
Dichotomous Scoring, Percent Agreement, and Cohen's Kappa*
for the Open-ended Mathematics Items Used in Cross-sectional Scaling
Age 9/Grade 4

Item	Block	Range of Response Codes	Correct Response Codes	Sample Size	Percent Agreement	Cohen's Kappa
N270901	MC	1-2	1	823	99	.95
N270001	MC	1	1	823	98	.97
N277601	MC	1-4	1	823	99	.97
N277602	MC	1-5	1	821	99	.97
N275401	MC	1-2	1	813	99	.97
N271101	MC	1-2	1	798	98	.96
M019701	ME	1-2	1	793	99	.99
M019801	ME	1-3	1-2	792	95	.91
M019901	ME	1-3	1-2	792	99	.98
M020001	ME	1-2	1	791	99	.99
M020101	ME	1-2	1	791	99	.97
M020201	ME	1-2	1	785	97	.96
M020301	ME	1-4	1	781	99	.98
M020401	ME	1-2	1	768	99	.98
M020501	ME	1-2	1	762	98	.97
N277903	ME	1-2	1	756	99	.97
M020701	ME	1-4	1	553	92	.80
M022201	MF	1-3	1-2	815	94	.88
M022501	MF	1-5	1	807	97	.93
M022801	MF	1-3	1	766	99	.98
M022802	MF	1-2	1	760	99	.99
M010631	MH	1-3	1	825	99	.97
M025531	MI	1-2	1	799	99	.98
M025831	MI	1-2	1	796	99	.98
M026431	MI	1-3	1	691	98	.96

* Cohen's Kappa is a measure of reliability that is appropriate for items that are dichotomized.

Table G-2 (continued)
Dichotomous Scoring, Percent Agreement, and Cohen's Kappa*
for the Open-ended Mathematics Items Used in Cross-sectional Scaling
Age 9/Grade 4

Item	Block	Range of Response Codes	Correct Response Codes	Sample Size	Percent Agreement	Cohen's Kappa
M026631	MI	1-2	1	608	99	.98
M026831	MI	1-2	1	513	99	.98
M026931	MI	1-2	1	455	96	.93
M034201	MK	1-2	1	748	96	.92
M034301	MK	1-2	1	748	99	.99
M034302	MK	1-2	1	748	99	.97
M034401	MK	1-3	1-2	748	96	.79
M034402	MK	1-2	1	748	99	.98
M034501	MK	1-2	1	748	98	.95
M034502	MK	1-2	1	748	99	.96
M036401	ML	1-2	1	748	99	.97
M036501	ML	1-4	1-2	748	99	.98
M036502	ML	1-2	1	748	99	.97
M036701	ML	1-2	1	748	99	.99
M036801	ML	1-4	1	748	99	.97
M036901	ML	1-3	1-2	748	97	.94

* Cohen's Kappa is a measure of reliability that is appropriate for items that are dichotomized.

Table G-3
Dichotomous Scoring, Percent Agreement, and Cohen's Kappa*
for the Open-ended Mathematics Items Used in Cross-sectional Scaling
Age 13/Grade 8

Item	Block	Range of Response Codes	Correct Response Codes	Sample Size	Percent Agreement	Cohen's Kappa
N276803	MC	1-2	1	794	99	.98
N277602	MC	1-5	1	794	99	.96
N256101	MC	1-2	1	793	99	.98
N286602	MC	1-3	1	791	100	.99
M019701	ME	1-2	1	809	100	1.00
M019801	ME	1-3	1-2	809	98	.96
M019901	ME	1-3	1-2	808	100	.99
M020001	ME	1-2	1	808	99	.99
M020101	ME	1-2	1	806	99	.99
M020201	ME	1-2	1	804	98	.93
M020301	ME	1-4	1	803	99	.98
M020401	ME	1-2	1	802	100	.99
M020501	ME	1-2	1	799	99	.98
M020801	ME	1-6	1	793	99	.99
M020901	ME	1-2	1	790	92	.87
M021001	ME	1-2	1	790	99	.98
M021101	ME	1-3	1	780	97	.94
M021201	ME	1-3	1	769	98	.96
M021301	ME	1-2	1	737	98	.95
M021302	ME	1-2	1	711	97	.94
M022201	MF	1-3	1-2	823	94	.88
M022501	MF	1-5	1	821	95	.91
M022801	MF	1-3	1	815	100	.99
M022802	MF	1-2	1	814	100	.99
M023701	MF	1-2	1	723	99	.98

* Cohen's Kappa is a measure of reliability that is appropriate for items that are dichotomized.

Table G-3 (continued)
Dichotomous Scoring, Percent Agreement, and Cohen's Kappa*
for the Open-ended Mathematics Items Used in Cross-sectional Scaling
Age 13/Grade 8

Item	Block	Range of Response Codes	Correct Response Codes	Sample Size	Percent Agreement	Cohen's Kappa
M016902	MG	1-2	1	765	97	.93
M013031	MH	1-4	1	808	99	.98
M013131	MH	1-3	1	798	99	.98
M027331	MI	1-3	1	775	99	.98
M027831	MI	1-2	1	768	99	.98
M027931	MI	1-2	1	764	99	.98
M028131	MI	1-3	1	738	99	.99
M028431	MI	1-3	1	684	98	.97
M028531	MI	1-2	1	659	96	.91
M028631	MI	1-5	1-2	577	98	.97
M036401	ML	1-2	1	770	98	.96
M036501	ML	1-4	1-2	770	98	.97
M036502	ML	1-2	1	770	95	.89
M036701	ML	1-2	1	770	99	.97
M036801	ML	1-4	1	770	99	.99
M036901	ML	1-3	1-2	770	97	.93
M037001	ML	1-3	1	770	98	.95

* Cohen's Kappa is a measure of reliability that is appropriate for items that are dichotomized.

Table G-4
Dichotomous Scoring, Percent Agreement, and Cohen's Kappa*
for the Open-ended Mathematics Items Used in Cross-sectional Scaling
Age 17/Grade 12

Item	Block	Range of Response Codes	Correct Response Codes	Sample Size	Percent Agreement	Cohen's Kappa
N260601	MC	1-2	1	796	98	.9[illegible]
N278501	MC	1-2	1	796	99	.98
N259001	MC	1-2	1	796	98	.97
N256001	MC	1-2	1	793	99	.98
N263001	MC	1-2	1	787	99	.99
M021401	ME	1-2	1	794	99	.98
M021501	ME	1-2	1	794	96	.92
M021502	ME	1-2	1	794	99	.99
M021601	ME	1-4	1	794	93	.80
M021602	ME	1-2	1	794	95	.89
M020201	ME	1-2	1	794	97	.85
M020301	ME	1-4	1	794	99	.98
M020401	ME	1-2	1	794	99	.99
M020501	ME	1-2	1	792	99	.98
M020801	ME	1-6	1	792	99	.98
M020901	ME	1-2	1	788	90	.85
M021001	ME	1-2	1	787	99	.98
M021101	ME	1-3	1	777	93	.86
M021201	ME	1-3	1	756	98	.96
M021701	ME	1-2	1	707	99	.98
M021702	ME	1-2	1	679	96	.94
M021801	ME	1-2	1	571	99	.95
M023901	MF	1-3	1	822	99	.99
M024701	MF	1-4	1-2	821	99	.98
M025302	MF	1-3	1	751	95	.87

* Cohen's Kappa is a measure of reliability that is appropriate for items that are dichotomized.

Table G-4 (continued)
Dichotomous Scoring, Percent Agreement, and Cohen's Kappa*
for the Open-ended Mathematics Items Used in Cross-sectional Scaling
Age 17/Grade 12

Item	Block	Range of Response Codes	Correct Response Codes	Sample Size	Percent Agreement	Cohen's Kappa
M016902	MG	1-2	1	791	94	.90
M017201	MG	1-3	1	696	99	.97
M017301	MG	1-2	1	563	98	.89
M013031	MH	1-4	1	804	98	.96
M013131	MH	1-3	1	798	99	.98
M011931	MH	1-2	1	576	98	.96
M012031	MH	1-3	1	547	98	.97
M029831	MI	1-3	1	793	98	.96
M030731	MI	1-3	1	662	97	.88
M030831	MI	1-4	1	604	98	.95
M036401	ML	1-2	1	802	98	.96
M037601	ML	1-2	1	802	99	.98
M037602	ML	1-4	1	802	99	.99
M037701	ML	1-4	1	802	98	.96
M037801	ML	1-4	1	802	99	.97
M037901	ML	1-2	1	802	99	.98
M038001	ML	1-2	1	802	97	.96
M037001	ML	1-3	1	802	97	.94
M037101	ML	1-4	1	802	98	.95
M037201	ML	1-3	1	802	97	.94
M037301	ML	1-4	1	802	99	.98
M037401	ML	1-3	1	802	99	.97
M037501	ML	1-3	1	802	100	.99

* Cohen's Kappa is a measure of reliability that is appropriate for items that are dichotomized.

Table G-5
Dichotomous Scoring, Percent Agreement, and Cohen's Kappa*
for the Open-ended Science Items Used in Cross-sectional Scaling
Age 9/Grade 4

Item	Block	Range of Response Codes	Correct Response Codes	Sample Size	Percent Agreement	Cohen's Kappa
K018701	SF	1-2	2	1017	95	.91
K018802	SF	1-3	2-3	897	89	.70
K018803	SF	1-5	2-5	805	89	.66
K025901	SH	1-4	4	985	95	.91
K026101	SH	1-6	6	970	93	.86
K026201	SH	1-5	5	967	85	.68
K026301	SH	1-6	6	960	99	.95
K026401	SH	1-4	4	946	98	.91
K026501	SH	1-4	4	934	97	.94
K026601	SH	1-5	5	901	95	.87
K026701	SH	1-5	5	869	97	.75
K026801	SH	1-4	4	846	95	.87
K026901	SH	1-5	5	714	96	.81
K028701	SI	1-4	4	840	98	.94
K028801	SI	1-2	2	715	95	.89
K028802	SI	1-4	4	670	99	.00

* Cohen's Kappa is a measure of reliability that is appropriate for items that are dichotomized.

Table G-6
Dichotomous Scoring, Percent Agreement, and Cohen's Kappa*
for the Open-ended Science Items Used in Cross-sectional Scaling
Age 13/Grade 8

Item	Block	Range of Response Codes	Correct Response Codes	Sample Size	Percent Agreement	Cohen's Kappa
K018701	SF	1-2	2	890	94	.88
K018801	SF	1-2	2	872	100	.98
K018802	SF	1-3	2-3	871	83	.67
K018803	SF	1-5	2-5	862	84	.71
K025501	SG	1-3	3	658	100	.99
K025502	SG	1-2	2	618	79	.56
K025503	SG	1-4	3-4	505	98	.75
K025901	SH	1-4	4	837	97	.92
K026101	SH	1-6	6	832	91	.82
K026201	SH	1-5	5	832	89	.77
K026301	SH	1-6	6	831	99	.98
K026401	SH	1-4	4	830	96	.92
K026501	SH	1-4	4	829	96	.93
K026601	SH	1-5	5	821	95	.89
K026701	SH	1-5	5	818	95	.68
K026801	SH	1-4	4	813	95	.90
K026901	SH	1-5	5	797	95	.91
K027101	SH	1-3	3	790	91	.82
K027201	SH	1-5	5	766	98	.95
K027301	SH	1-4	4	734	98	.94
K027401	SH	1-6	6	706	98	.93
K028701	ST	1-4	4	884	96	.88
K028801	SI	1-2	2	879	95	.91
K028802	SI	1-4	4	879	98	.95

* Cohen's Kappa is a measure of reliability that is appropriate for items that are dichotomized.

Table G-7
Dichotomous Scoring, Percent Agreement, and Cohen's Kappa*
for the Open-ended Science Items Used in Cross-sectional Scaling
Age 17/Grade 12

Item	Block	Range of Response Codes	Correct Response Codes	Sample Size	Percent Agreement	Cohen's Kappa
K021301	SF	1-4	2-4	574	83	.65
K025501	SG	1-3	3	702	100	.99
K025502	SG	1-2	2	665	83	.63
K025503	SG	1-4	3-4	580	97	.94
K025901	SH	1-4	4	825	96	.87
K026101	SH	1-6	6	824	89	.77
K026201	SH	1-5	5	823	87	.74
K026301	SH	1-6	6	820	98	.96
K026401	SH	1-4	4	817	97	.94
K026501	SH	1-4	4	816	97	.94
K026601	SH	1-5	5	812	96	.92
K026701	SH	1-5	5	811	94	.82
K026801	SH	1-4	4	808	95	.91
K026901	SH	1-5	5	800	95	.90
K027101	SH	1-3	3	790	89	.71
K027201	SH	1-5	5	756	96	.92
K027301	SH	1-4	4	739	99	.96
K027401	SH	1-6	6	719	99	.98
K027501	SH	1-4	3-4	559	90	.77
K027601	SH	1-6	6	425	96	.88
K030601	SI	1-2	2	799	96	.89
K030602	SI	1-3	3	788	98	.95
K030603	SI	1-3	2-3	740	86	.72
K030701	SI	1-4	2-4	574	81	.62

* Cohen's Kappa is a measure of reliability that is appropriate for items that are dichotomized.

APPENDIX H

Differential Item Functioning (DIF) Results

Table H-1
Reading Items Identified as "C" Items in at least One Comparison*

Item	Block	Grade	Comparison	Group Favored
R010601	RG	4,12	White/Black	White
R011803	RI	12	Male/Female	Male
R010104	RD	12	Male/Female	Female
N001601	RD	8	White/Hispanic	White
N004202	RF	12	White/Black	Black
R000904	RE	4	Male/Female White/Black	Male White
R001501	RE	12	Male/Female White/Black	Female White
R001503	RE	8	Male/Female White/Black	Female Black
R001504	RE	12	White/Black	White
R001604	RE	8	White/Black	Black

* For each grade for which an item was administered, three comparisons were performed: Male/Female, White/Black, and White/Hispanic.

Table H-2
Mathematics Items Identified as "C" Items in at least One Comparison*

Item	Block	Subscale/Item Type	Grade	Comparison	Groups Favored
M017801	MD	Data Analysis, Statistics, & Probability	12	White/Black	White
M016201	MG	Measurement	12	Male/Female	Male
M029531	MI	Data Analysis, Statistics, & Probability	12	White/Black	White
M019701	ME	Algebra and Functions	4	Male/Female	Female
M022501	MF	Geometry	4	White/Black	Black
M022201	MF	Geometry	4	White/Black	White
M017201	MG	Geometry	12	White/Black	White
M034702	MK	Estimation	8	Male/Female	Female
M034801	MK	Estimation	12	Male/Female	Female
M036601	ML	Higher-order Thinking Skills	4	White/Black	White
M036901	ML	Higher-order Thinking Skills	8	White/Black	Black
M013631	MH	Numbers and Operations	8	Male/Female	Male
M016401	MG	Geometry	12	White/Black	White
M013531	MH	Numbers and Operations	8	White/Black	White
M032001	MJ	Estimation	4	White/Black	White
M036701	ML	Higher-order Thinking Skills	4 8	Male/Female White/Black	Female Black
M020301	ME	Measurement	8 12	White/Black Male/Female White/Black	White Male White
M020501	ME	Numbers and Operations	12	Male/Female White/Black White/Hispanic	Male White White
M028731	MI	Numbers and Operations	8	Male/Female	Male
M036401	ML	Higher-order Thinking Skills	4 12	White/Black White/Black	Black Black
M038001	ML	Higher-order Thinking Skills	12	White/Black	Black
N255701	MC	Algebra and Functions	12	Male/Female	Female
N266501	MC	Measurement	12	Male/Female	Male

* For each grade for which an item was administered, three comparisons were performed: Male/Female, White/Black, and White/Hispanic.

Table H-3

Science Items Identified as "C" Items in at least One Comparison*

Item	Block	Subscale	Grade	Comparison	Group Favored
N400201	SC	Life Sciences	4,8,12	Male/Female	Female
N403001	SC	Nature of Science	4	Male/Female	Female
N404701	SC	Life Sciences	8	Male/Female	Female
N405501	SC	Earth & Space Sciences	12	Male/Female	Male
N407201	SC	Physical Sciences	12	Male/Female	Male
N410101	SC	Nature of Science	12	Male/Female	Female
N411501	SC	Life Sciences	12	Male/Female	Female
N411801	SC	Physical Sciences	12	Male/Female	Female
K010101	SD	Earth & Space Sciences	4	White/Black	White
K010301	SD	Nature of Science	12	Male/Female	Female
K010501	SD	Physical Sciences	8,12	Male/Female	Female
K011101	SD	Earth & Space Sciences	12	White/Hispanic	White
K011301	SD	Nature of Science	8,12	Male/Female	Female
K012201	SD	Physical Sciences	12	Male/Female	Male
K012901	SD	Nature of Science	12	Male/Female	Female
K013101	SE	Physical Science	8	White/Black	Black
K013301	SE	Earth & Space Sciences	12	Male/Female	Male
K013701	SE	Life Sciences	8	Male/Female	Male
K014201	SE	Physical Sciences	8	White/Black	Black
K016001	SE	Nature of Science	12	Male/Female	Female
K018701	SF	Earth & Space Sciences	4,8	White/Black	White
K018803	SF	Physical Sciences	4,8	Male/Female	Female
K019701	SF	Life Sciences	12	White/Black	White
K012901	SG	Life Sciences	4	White/Black	Black
K021601	SG	Physical Sciences	4	Male/Female	Female
K024201	SG	Earth & Space Sciences	12	White/Black	Black

* For each grade for which an item was administered, three comparisons were performed: Male/Female, White/Black, and White/Hispanic.

Table H-3 (continued)

Science Items Identified as "C" Items in at least One Comparison*

Item	Block	Subscale	Grade	Comparison	Group Favored
K025503	SG	Physical Sciences	12	Male/Female	Male
K026601	SH	Life Sciences	4	Male/Female	Female
K027101	SH	Earth & Space Sciences	8	Male/Female	Female
K027901	SI	Life Sciences	8	Male/Female	Male
K028301	SI	Physical Sciences	8	Male/Female	Male
K028701	SI	Physical Sciences	4,8	Male/Female	Male
K028801	SI	Physical Sciences	8	White/Black	White
K029601	SI	Nature of Science	12	Male/Female	Female

* For each grade for which an item was administered, three comparisons were performed: Male/Female, White/Black, and White/Hispanic.

GLOSSARY OF TERMS

Glossary of Terms

anchoring. The process of characterizing score levels in terms of predicted observable behavior.

assessment session. The period of time during which a NAEP booklet is administered to one or more individuals.

average response method (ARM). A regression-based technique to predict for a respondent the conditional distribution of an average score on a set of items given responses to at least one of the items and other information.

background questionnaires. The instruments used to collect information about students' demographics and educational experiences.

bias. In statistics, the difference between the expected value of an estimator and the population parameter being estimated. If the average value of the estimator over all possible samples (the estimator's expected value) equals the parameter being estimated, the estimator is said to be **unbiased**; otherwise, the estimator is **biased**.

BIB (Balanced Incomplete Block) spiraling. A complex variant of multiple matrix sampling, in which items are administered in such a way that each pair of items is administered to a nationally representative sample of respondents.

BILOG. A computer program for estimating item parameters.

block. A group of assessment items created by dividing the item pool for an age/grade into subsets. Used in the implementation of the BIB spiral sample design.

booklet. The assessment instrument created by combining blocks of assessment items.

bridging. An administration of the same set of exercises under two different conditions or to two different populations to allow a statistical link (**bridge**) to be established between results under the different circumstances.

calibrate. To estimate the parameters of a set of items from responses of a sample of examinees.

clustering. The process of forming sampling units as groups of other units.

codebook. A formatted printout of NAEP data for each student, excluded student, teacher, and school in a particular grade/age.

coefficient of variation. The ratio of the standard deviation of an estimate to the value of the estimate.

combined ratio estimator. The ratio estimator resulting from first estimating the numerator and the denominator values and then using the quotient of these as the estimate of the ratio.

common block. A group of background items included in the beginning of every assessment booklet.

conditional probability. Probability of an event, given the occurrence of another event.

conditioning variables. Demographic and other background variables characterizing a respondent. Used in construction of plausible values.

cross-sectional assessment. An assessment that provides information about differences in educational performance across subgroups of students. It does not provide information about changes in students' educational

performance across time. It may, however, provide baseline data for measuring future trends.

Current Population Survey. A household sample survey conducted monthly by the Bureau of the Census to provide estimates of employment, unemployment, and other characteristics of the general labor force, the population as a whole, and various subgroups of the population.

degrees of freedom. [of a variance estimator] The number of independent pieces of information used to generate a variance estimate.

derived variables. Subgroup data that were not obtained directly from assessment responses, but through procedures of interpretation, classification, or calculation.

design effects. The ratio of the variance for the sample design to the variance for a simple random sample of the same size.

distractor. An incorrect response choice included in a multiple-choice item.

excluded student questionnaire. An instrument completed for every student who was sampled but excluded from the assessment.

excluded students. Sampled students determined by the school to be unable to participate because they have limited English proficiency, are mildly mentally retarded (educable), or are functionally disabled.

expected value. The average of the sample estimates given by an estimator over all possible samples. If the estimator is unbiased, then its expected value will equal the population value being estimated.

field test. A pretest of items to obtain information regarding clarity, difficulty levels, timing, feasibility, and special administrative situations; performed before revising and selecting items to be used in the assessment.

focused-BIB spiraling. A variation of BIB spiraling in which items are administered in such a way that each pair of items *within a subject area* is administered to a nationally representative sample of respondents.

foils. The correct and incorrect response choices included in a multiple-choice item.

group effect. The difference between the mean for a group and the mean for the nation.

holistic scoring. A method of evaluating students' writing for overall fluency in responding to a task.

imputation. Prediction of a missing value according to some procedure, using a mathematical model in combination with available information. See **plausible values.**

imputed race/ethnicity. The race or ethnicity of an assessed student, as derived from his or her responses to particular common background items. A NAEP **reporting subgroup.**

item response theory (IRT). Test analysis procedures that assume a mathematical model for the probability that a given examinee will respond correctly to a given exercise.

jackknife. A procedure to estimate standard errors of percentages and other statistics. Particularly suited to complex sample designs.

machine-readable catalog. Computer processing control information, IRT parameters, foil codes, and labels in a computer-readable format.

major strata. Used to stratify the primary sampling frame within each region. Involves stratification by size of community and degree of ruralization (SDOC).

metropolitan statistical area (MSA). An area defined by the federal government for the purposes of presenting general-purpose statistics for metropolitan areas. Typically, an MSA contains a city with a population of at

least 50,000 plus adjacent areas.

modal age. The age of the majority of a group of grade-eligible students.

modal grade. The grade attended by the majority of a group of age-eligible students.

mode of administration. The method by which students are administered assessment instruments. Both printed and tape-recorded administration methods are used.

multistage sample design. Indicates more than one stage of sampling. An example of three-stage sampling: 1) sample of counties (primary sampling units or PSUs); 2) sample of schools within each sample county; 3) sample of students within each sample school.

multiple matrix sampling. Sampling plan in which different samples of respondents take different samples of items.

NAEP scales. The anchored scales common across age/grade levels and assessment years used to report NAEP results.

nonresponse. The failure to obtain responses or measurements for all sample elements.

nonsampling error. A general term applying to all sources of error except sampling error. Includes errors from defects in the sampling frame, response or measurement error, and mistakes in processing the data.

objective. A desirable education goal agreed upon by scholars in the field, educators, and concerned laypersons, and established through the consensus approach.

observed race/ethnicity. Race or ethnicity of an assessed student as perceived by the exercise administrator.

open-ended response item. A nonmultiple-choice item that requires some type of written or oral response.

oversampling. Deliberately sampling a portion of the population at a higher rate than the remainder of the population.

paced tape. The audio recording that accompanies some booklets to assure uniformity in administration. Recorded instructions prevent reading difficulties from interfering with an individual's ability to respond.

parental education level. The level of education of the mother and father of an assessed student as derived from the student's response to two assessment items. A NAEP **reporting subgroup.**

percent correct. The percent of a target population that would answering a particular exercise correctly.

plausible values. Proficiency values drawn at random from a conditional distribution of a NAEP respondent, given his or her response to cognitive exercises and a specified subset of background variables (conditioning variables). The selection of a plausible value is a form of **imputation.**

poststratification. Classification and weighting to correspond to external values of selected sampling units by a set of strata definitions after the sample has been selected.

primary sampling unit (PSU). The basic geographic sampling unit for NAEP. Either a single county or a set of contiguous counties.

primary trait scoring. A method of evaluating students' writing for effectiveness in accomplishing the specific goal or purpose of each writing task.

principal questionnaire. A data collection form given to school principals before assessments. The principals respond to questions concerning enrollment, size and occupational composition of the community, etc.

probability sample. A sample in which every element of the population has a known,

nonzero probability of being selected.

pseudoreplicate. The value of a statistic based on an altered sample. Used by the **jackknife** variance estimator.

QED. Quality Education Data, Inc. A supplier of lists of schools, school districts, and other school data.

random variable. A variable that takes on any value of a specified set with a particular probability.

region. One of four geographic areas used in gathering and reporting data: Northeast, Southeast, Central, and West (as defined by the Office of Business Economics, U.S. Department of Commerce). A NAEP **reporting subgroup.**

reporting subgroup. Groups within the national population for which NAEP data are reported: for example, gender, race/ethnicity, grade, age, level of parental education, region, and type of community.

respondent. A person who is eligible for NAEP, is in the sample, and responds by completing one or more items in an assessment booklet.

response options. In a multiple-choice question, alternatives that can be selected by a respondent.

sample. A portion of a population, or a subset from a set of units, selected by some probability mechanism for the purpose of investigating the properties of the population. NAEP does not assess an entire population but rather selects a representative sample from the group to answer assessment items.

sampling error. The error in survey estimates that occurs because only a sample of the population is observed. Measured by sampling **standard error.**

sampling frame. The list of sampling units from which the sample is selected.

sampling weight. A multiplicative factor equal to the reciprocal of the probability of a respondent being selected for assessment with adjustment for nonresponse and perhaps also for poststratification. The sum of the weights provides an estimate of the number of persons in the population represented by a respondent in the sample.

school characteristics and policy questionnaire. A questionnaire completed for each school by the principal or other official; used to gather information concerning school administration, staffing patterns, curriculum, and student services.

secondary-use data files. Computer files containing respondent-level cognitive, background and attitude, and demographic data. Available from NCES for use by researchers wishing to do secondary analyses of NAEP data.

selection probability. The chance that a particular sampling unit has of being selected in the sample.

session. A group of students reporting for the administration of an assessment.

simple random sample. Process for selecting n sampling units from a population of N sampling units so that each sampling unit has an equal chance of being in the sample and every combination of n sampling units has the same chance of being in the sample chosen.

standard error. A measure of sampling variability and measurement error for a statistic. Because of NAEP's complex sample design, sampling standard errors are estimated by **jackknifing** the samples from first-stage sample estimates. Standard errors may also include a component due to the error of measurement of individual scores estimated using plausible values.

stratification. The division of a population into parts, called strata.

stratified sample. A sample selected from a

population that has been stratified, with a sample selected independently in each stratum. The strata are defined for the purpose of reducing sampling error.

student ID number. A unique identification number assigned to each respondent to preserve his or her anonymity. NAEP does not record the names of any respondents.

subject area. One of the areas assessed by National Assessment; for example, art, civics, computer competence, geography, literature, mathematics, music, reading, science, U.S. history, or writing.

systematic sample (systematic random sample). A sample selected by a systematic method; for example, when units are selected from a list at equally spaced intervals.

teacher questionnaire. A questionnaire completed by selected teachers of sample students; used to gather information concerning years of teaching experience, frequency of assignments, teaching materials used, and availability and use of computers.

trend assessment. An assessment based on replicating past procedures in order to report changes in educational achievement across time.

trimming. A process by which extreme weights are reduced (trimmed) to diminish the effect of extreme values on estimates and estimated variances.

type of community (TOC). One of the NAEP **reporting subgroups**, dividing the communities in the nation into four groups on the basis of size and other characteristics.

variance. The average of the squared deviations of a random variable from the expected value of the variable. The variance of an estimate is the squared standard error of the estimate.

REFERENCES CITED IN TEXT

References Cited in Text

Beaton, A. E. (1987). The NAEP reading scale. In A. E. Beaton, *Implementing the new design: The NAEP 1983-84 technical report* (pp. 381-390). (No. 15-TR-20) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Beaton, A. E. (1988). *Expanding the new design: The NAEP 1985-86 technical report.* (No. 17-TR-20) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Beaton, A. E. & Allen, N. (in press). Interpreting scales through scale anchoring. *Journal of Educational Statistics.*

Beaton, A. E., & Johnson, E. G. (1990). The average response method of scaling. *Journal of Educational Statistics, 15,* 9-38..

Beaton, A. E., & Johnson, E. G. (1987). The average response method (ARM) of scaling. In A. E. Beaton, *Implementing the new design: The NAEP 1983-84 technical report* (pp. 435-480). (No. 15-TR-20) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Beaton, A. E., & Zwick, R. (1990). *The effect of changes in the National Assessment: Disentangling the NAEP 1985-86 reading anomaly.* (No. 17-TR-21) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: Theory and practice.* Cambridge, MA: MIT Press.

Braun, H. I. (1988). Understanding scoring reliability: Experiments in calibrating essay readers. *Journal of Educational Statistics. 13*(1), 1-18.

Caldwell, N., Slobasky, R., Moore, D., & Ter Maat, J. (1991). *Report on field operations and data collection activities, 1990 National Assessment of Educational Progress.* Rockville, MD: Westat, Inc.

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). New York: John Wiley & Sons.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin, 70*(4), 213-220.

Dossey, J. A., Mullis, I. V. S., Lindquist, M. M., & Chambers, D. L. (1988). *The mathematics report card: Are we measuring up? Trends and achievement based on the 1986 National Assessment.* Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Educational Testing Service (1987). *ETS standards for quality and fairness.* Princeton, NJ: Author.

Foertsch, M. A. (1992). *Reading in school and out of school: Students' literary experience and academic achievement from 1988 to 1990 at grades 4, 8, and 12.* Washington, DC: National Center for Education Statistics.

Gentile, C. A. (1992). *The writing students do in school: The 1990 NAEP portfolio study of fourth and eighth graders' school-based writing.* Washington, DC: National Center for Education Statistics.

Hambleton, R. K. (1989). Principles and selected applications of item response theory. In R. L. Linn (Ed.), *Educational measurement* (3rd edition). New York: American Council on Education/Macmillan.

Holland, P. W. (1985). On the study of differential item performance without IRT. *Proceedings of the Military Testing Association*, October.

Holland, P. W. & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity*. Hillsdale, NJ: Erlbaum.

Illinois State Board of Education. (1988). *Assessing reading in Illinois.* Springfield, IL: Author.

Johnson, E. G. (1988). Mathematics data analysis. In A. E. Beaton, *Expanding the new design: The NAEP 1985-86 technical report* (pp. 215-240). (No. 17-TR-20) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Johnson, E. G. (1989). Considerations and techniques for the analysis of NAEP data. *Journal of Educational Statistics, 14*, 303-334.

Johnson, E. G. (1990). Data analysis for the writing assessment. In E. G. Johnson & R. Zwick, *Focusing the new design: The NAEP 1988 technical report* (pp. 267-296). (No. 19-TR-20) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Johnson, E. G. & King, B. F. (1987). Generalized variance functions for a complex sample survey. *Journal of Official Statistics, 3*, 235-250.

Johnson, E. G., & Rust, K. F. (in press). Population inferences and variance estimation for NAEP data. *Journal of Educational Statistics.*

Johnson, E. G., Rust, K. F., & Hansen, M. H. (1990). Weighting procedures and estimation of sampling variance. In E. G. Johnson & R. Zwick, *Focusing the new design: The NAEP 1988 technical report* (pp. 187-228). (No. 19-TR-20) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Johnson, E. G. & Zwick, R. (1990). *Focusing the new design: The NAEP 1988 technical report.* (No. 19-TR-20) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Johnson, J. R. (1987). Instrument and item information. In A. E. Beaton, *Implementing the new design: The NAEP 1983-84 technical report* (pp. 119-134). (No. 15-TR-20) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Jones, L. R., Mullis, I. V. S., Raizen, S. A., Weiss, I. R., & Weston, E. A. (1992). The 1990 science report card: NAEP's assessment of fourth, eighth, and twelfth graders. Washington, DC: National Center for Education Statistics.

Joreskog, K. G. & Sorbom, D. (1986). *PRELIS: A program for multivariate data screening and data summarization. A preprocessor for LISREL.* [Computer program]. Mooresville, IN: Scientific Software.

Kaplan, B. A., Beaton, A. E., Johnson, E. G., & Johnson, J. R. (1988). *National Assessment of Educational Progress 1986 bridge studies.* (No. 17-TR-11) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Kennel, T., Mohn, L., Reynolds, L., Smrdel, D., Thayer, B., & Zaback, L. (1991). *Report of processing and professional scoring activities, NAEP - 1989-90.* Iowa City, IA: National Computer Systems.

Kish, L. (1965). *Survey sampling.* New York: John Wiley & Sons.

Kish, L. & Frankel, M. R. (1974). Inference from complex samples. *Journal of the Royal Statistical Society, Series B, 36*, 1-22.

Koffler, S. L. (1991). *The technical report of NAEP's 1990 Trial State Assessment.* (No. 21-ST-01) Washington, DC: National Center for Education Statistics.

Krewski, D. & Rao, J. N. K. (1981). Inference from stratified samples: Properties of linearization, jackknife and balanced repeated replication. *Annals of Statistics, 9*, 1010-1019.

Langer, J. A., Applebee, A. N., Mullis, I. V. S., & Foertsch, M. A. (1990). *Learning to read in our nation's schools: Instruction and achievement in 1988 at grades 4, 8, and 12.* Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Little, R. J. A., & Rubin, D. B. (1983). On jointly estimating parameters and missing data. *American Statistician, 37*, 218-220.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Michigan State Board of Education. (1987). *Blueprint for the new MEAP reading test.* Lansing, MI: Author.

Mislevy, R. J. (1984). Estimating latent distributions. *Psychometrika, 49*(3), 359-381.

Mislevy, R. J. (1985). Estimation of latent group effects. *Journal of the American Statistical Association, 80*, 993-997.

Mislevy, R. J. (1991). Randomization-based inferences about latent variables from complex samples. *Psychometrika, 56*(2), 177-196.

Mislevy, R. J., & Bock, R. D. (1982). *BILOG: Item analysis and test scoring with binary logistic models* [Computer program]. Mooresville, IN: Scientific Software.

Mislevy, R. J., & Bock, R. D. (1989). *PC-BILOG 3: Item analysis and test scoring with binary logistic models.* Mooresville, IN: Scientific Software.

Mislevy, R. J., Johnson, E. G., & Muraki, E. (in press). Scaling procedures in NAEP. *Journal of Educational Statistics.*

Mislevy, R. J., & Sheehan, K. M. (1987). Marginal estimation procedures. In A. E. Beaton, *Implementing the new design: The NAEP 1983-84 technical report* (pp. 293-360). (No. 15-TR-20) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Mislevy, R. J., & Sheehan, K. M. (1989). The role of collateral information about examinees in item parameter estimation. *Psychometrika, 54*, 661-679.

Mislevy, R. J., & Wu, P-K. (1988). *Inferring examinee ability when some item responses are missing.* (ETS Research Report RR-88-48-ONR) Princeton, NJ: Educational Testing Service.

Mullis, I. V. S. (1990). *The NAEP guide.* Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Mullis, I. V. S., Dossey, J. A., Foertsch, M. A., Jones, L. R., & Gentile, C. A. (1991). *Trends in academic progress: Achievement of U.S. students in science, 1969-70 to 1990; mathematics, 1973 to 1990; reading, 1971 to 1990; and writing, 1984 to 1990.* Washington, DC: National Center for Education Statistics.

Mullis, I. V. S., Dossey, J. A., Owen, E. H., & Phillips, G. W. (1991). *The state of mathematics achievement: NAEP's 1990 assessment of the nation and the trial assessment of the states.* Washington, DC: National Center for Education Statistics.

Muthén, B. (1991). *Issues in using NAEP mathematics items to study achievement dimensionality, within-grade differences, and across-grade growth.* Report presented to the Design and Analysis Committee for NAEP.

National Assessment of Educational Progress. (1985a). *The reading report card: Progress toward excellence in our schools.* (No. 15-R-01) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

National Assessment of Educational Progress. (1985b). *Science objectives: 1985-86 assessment.* Princeton, NJ: Educational Testing Service.

National Assessment of Educational Progress (1987). *Mathematics objectives: 1985-86 assessment.* Princeton, NJ: Educational Testing Service.

National Assessment of Educational Progress (1988). *Mathematics objectives, 1990 assessment.* Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

National Assessment of Educational Progress (1989a). *Reading objectives, 1990 assessment.* Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

National Assessment of Educational Progress (1989b). *Science objectives, 1990 assessment.* Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

National Assessment of Educational Progress (1991a). *The 1990 reading cross-sectional report.* Princeton, NJ: Educational Testing Service.

National Assessment of Educational Progress (1991b). *The 1990 science cross-sectional report.* Princeton, NJ: Educational Testing Service.

National Assessment of Educational Progress (1991c). *The 1990 writing portfolio report.* Princeton, NJ: Educational Testing Service.

National Council of Teachers of Mathematics (1987). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

Raizen, S., & Jones, L. (Eds.) (1985). *Indicators of precollege education in science and mathematics: A preliminary review*. Washington, DC: National Academy Press.

Rogers, A. M. (1991). *NAEP-MGROUP: Enhanced version of Sheehan's software for the estimation of group effects in multivariate models* [Computer program]. Princeton, NJ: Educational Testing Service.

Rogers, A. M., Barone, J. L., & Kline, D. L. (1990). *A guide to the NAEP item information database*. Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Rogers, A. M., Barone, J. L., & Kline, D. L. (1989). *A guide to the NAEP restricted-use data files*. Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Rogers, A. M., Kline, D. L., Barone, J. L., Mychajlowycz, A. W., & Forer, D. C. (1989). *A primer for the NAEP item information database*. Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Rogers, A. M., Kline, D. L., Johnson, E. G., Mislevy, R. M., Allen, N. A., & Rust, K. F. (1992). *National Assessment of Educational Progress 1990 secondary-use data files user guide*. Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Rogers, A. M., & Nelson, J. E. (1990). *NAEP-BILOG: Enhanced version of Mislevy and Bock's software for item analysis and test scoring with binary logistic models* [Computer program]. Princeton, NJ: Educational Testing Service.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons.

Rust, K. (1985). Variance estimation for complex estimators in sample surveys. *Journal of Official Statistics, 1(4)*, 381-397.

Rust, K. F., Bethel, J., Burke, J., & Hansen, M. H. (1990). *The 1988 National Assessment of Educational Progress--Sampling and weighting procedures, final report*. Rockville, MD: Westat, Inc.

Rust, K. F., Burke, J., & Fahimi, M. (1992). *1990 National Assessment of Educational Progress sampling and weighting procedures, part 2: National assessment, final report*. Rockville, MD: Westat, Inc.

Satterthwaite, F. E. (1941). Synthesis of variance. *Psychometrika*, *6*, 309-316.

Sheehan, K. M. (1985). *MGROUP: Estimation of group effects in multivariate models* [Computer program]. Princeton, NJ: Educational Testing Service.

Tobias, S. (1987). *Succeed with math: Every student's guide to conquering mathematics anxiety*. New York: College Entrance Examination Board.

Westat, Inc. (1984). *The private school survey. Final report*. Submitted to the National Center for Education Statistics, U. S. Department of Education. Rockville, MD: Author.

Wingersky, M., Kaplan, B. A., & Beaton, A. E. (1987). Joint estimation procedures. In A. E. Beaton, *Implementing the new design: The NAEP 1983-84 technical report* (pp. 285-292). (No. 15-TR-20) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Wolter, K. M. (1985). *Introduction to variance estimation.* New York: John Wiley & Sons.

Yamamoto, K. (1988). Science data analysis. In A. E. Beaton, *Expanding the new design: The NAEP 1985-86 technical report* (pp. 243-255). (No. 17-TR-20) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Yamamoto, K. (1990). Mathematics and science trend data analysis. In A. E. Beaton & R. Zwick, *The effect of changes in the National Assessment: Disentangling the NAEP 1985-86 reading anomaly* (pp. 111-144). Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Zwick, R. (1987a). Assessing the dimensionality of NAEP reading data. *Journal of Educational Measurement, 24*, 293-308.

Zwick, R. (1987b). Use of the NAEP almanacs. In A. E. Beaton, *Implementing the new design: The NAEP 1983-84 technical report* (pp. 517-522). (No. 15-TR-20) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Zwick, R. (1988). *Professionally scored items.* Technical memorandum.

Zwick, R. & Ercikan, K. (1989). Analysis of differential item functioning in the NAEP history assessment. *Journal of Educational Measurement, 26*, 55-66.

Zwick, R. & Grima, A. (1991). *Policy for differential item functioning (DIF) analysis in NAEP.* Technical memorandum.

INDEX

Index

838

DEPARTMENT OF EDUCATION

UNITED STATES OF AMERICA

91232-05819 • S32M.20 • Printed in U.S.A.